

Yale University

## EliScholar – A Digital Platform for Scholarly Publishing at Yale

---

Yale Graduate School of Arts and Sciences Dissertations

---

Spring 2021

### Essays on the Effects of Institutional Changes

Martin Mattsson

Yale University Graduate School of Arts and Sciences, [martin.mattsson@yale.edu](mailto:martin.mattsson@yale.edu)

Follow this and additional works at: [https://elischolar.library.yale.edu/gsas\\_dissertations](https://elischolar.library.yale.edu/gsas_dissertations)

---

#### Recommended Citation

Mattsson, Martin, "Essays on the Effects of Institutional Changes" (2021). *Yale Graduate School of Arts and Sciences Dissertations*. 87.

[https://elischolar.library.yale.edu/gsas\\_dissertations/87](https://elischolar.library.yale.edu/gsas_dissertations/87)

This Dissertation is brought to you for free and open access by EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Yale Graduate School of Arts and Sciences Dissertations by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact [elischolar@yale.edu](mailto:elischolar@yale.edu).

## Abstract

# Essays on the Effects of Institutional Changes

Martin Mattsson

2021

Formal and informal institutions are important determinants of behavior and economic outcomes. In this dissertation, I study the causal effects of changes to one formal and one informal institution on individuals' behavior. The first chapter uses an experiment in Bangladesh to show how providing information on delays in a public service provision to government bureaucrats and their supervisors affects these bureaucrats' behavior and the outcomes for applicants for the public service. The second chapter (co-authored with Ro'ee Levy) shows how the MeToo movement increased the reporting of sexual crimes to the police by changing the norms, information, or both, about sexual misconduct.

### *Chapter 1: "Service Delivery, Corruption, and Information Flows in Bureaucracies: Evidence from the Bangladesh Civil Service"*

Government bureaucracies in low- and middle-income countries often suffer from corruption and slow public service delivery. Can an information system – providing information about delays to the responsible bureaucrats and their supervisors – reduce delays? Paying bribes for faster service delivery is a common form of corruption, but does improving average processing times reduce bribes? To answer these questions, I conduct a large-scale field experiment over 16 months with the Bangladesh Civil Service. I send monthly scorecards measuring delays in service delivery to government officials and their supervisors. The scorecards increase services delivered on time by 11% but do not reduce bribes. Instead, the scorecards *increase* bribes for high-performing bureaucrats. These results are inconsistent with existing theories suggesting

that speeding up service delivery reduces bribes. I propose a model where bureaucrats' shame or reputational concerns constrain corruption. When bureaucrats' reputation improves through positive performance feedback, this constraint is relaxed, and bribes increase. Overall, my study shows that improving information within bureaucracies can change bureaucrats' behavior, even without explicit incentives. However, positive performance feedback can have negative spillovers on bureaucrats' performance across different behaviors.

*Chapter 2: "The Effects of Social Movements: Evidence from #MeToo" (Joint with Ro'ee Levy)*

Social movements are associated with large societal changes, but evidence on their causal effects is limited. We study the effect of the MeToo movement on a high-stakes decision—reporting a sexual crime to the police. We construct a new dataset of sexual and non-sexual crimes reported in 30 OECD countries, covering 88% of the OECD population. We analyze the effect of the MeToo movement by employing a triple-difference strategy over time, across countries, and between crime types. The movement increased reporting of sexual crimes by 10% during its first six months. The effect is persistent and lasts at least 15 months. Because we find a strong effect on reporting before any major changes to laws or policy took place, we attribute the effect to a change in social norms or information. Using more detailed US data, we show that the movement also increased arrests for sexual crimes in the long run. In contrast to a common criticism of the movement, we do not find evidence for large differences in the effect across racial and socioeconomic groups. Our results suggest that social movements can rapidly change high-stakes personal decisions.

# Essays on the Effects of Institutional Changes

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

in Candidacy for the Degree of

Doctor of Philosophy

by

Martin Mattsson

Dissertation Director: Ahmed Mushfiq Mobarak

June 2021

©2021 by Martin Mattsson

All rights reserved.

# Contents

<b>I Service Delivery, Corruption, and Information Flows in Bureaucracies: Evidence from the Bangladesh Civil Service</b>	<b>1</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Context, Experimental Intervention, and Data</b>	<b>6</b>
2.1 Land Record Changes in Bangladesh . . . . .	7
2.2 E-governance System for Land Record Changes . . . . .	9
2.3 Experimental Intervention: Performance Scorecards . . . . .	9
2.4 Randomization . . . . .	11
2.5 Data . . . . .	11
2.6 Balance of Randomization . . . . .	14
2.7 Additional Intervention: Providing Information to Applicants . . . . .	15
<b>3 Empirical Strategy</b>	<b>15</b>
3.1 Empirical Strategy: Overall Effects . . . . .	15
3.2 Empirical Strategy: Heterogeneous Effects . . . . .	16
3.3 Analysis of Additional Experiments and Potential Interactions . . . . .	17
<b>4 Results: Effects on Processing Times and Bribes</b>	<b>17</b>
4.1 Effect on Processing Times . . . . .	18
4.2 Mechanisms for the Effect on Processing Times . . . . .	20
4.3 Effect on Visits to Land Office and Time Spent by Applicants . . . . .	21
4.4 Effect on Bribe Payments . . . . .	22
4.5 Heterogeneity of Results by Office Performance at Baseline . . . . .	23
4.6 External Validity, Potential Biases from Surveying, and Unintended Consequences <sup>1</sup> . . . . .	24
<b>5 Implications for Theories of the Relationship between Processing Times and Bribes</b>	<b>25</b>
5.1 Do Faster Processing Times Decrease Bribe Payments? . . . . .	27

---

<sup>1</sup>Appendix Section C discusses potential biases, external validity, and unintended consequences in more detail.

5.2	Implications for Other Theories . . . . .	28
<b>6</b>	<b>Model of Bureaucrat Behavior</b>	<b>28</b>
6.1	Model Set-up . . . . .	29
6.2	Model Predictions . . . . .	30
6.3	Potential General Equilibrium Effects within the Civil Service . . . . .	32
6.4	Alternative Explanations for the Effects of Scorecards . . . . .	33
<b>7</b>	<b>Conclusion</b>	<b>35</b>
	<b>Appendices</b>	<b>50</b>
<b>A</b>	<b>Theory</b>	<b>50</b>
A.1	Model of Reputation and Bureaucrat Behavior . . . . .	50
A.2	Monopolistic Price Discrimination Model . . . . .	53
<b>B</b>	<b>Additional Details on Experiment and Data</b>	<b>57</b>
B.1	Details on Randomization of Scorecards Treatment Assignment . . . . .	57
B.2	Data . . . . .	58
<b>C</b>	<b>Additional Empirical Analysis</b>	<b>62</b>
C.1	External Validity of Results . . . . .	62
C.2	Potential Bias from Applicant Survey and Information Intervention . . . . .	64
C.3	Unintended Consequences of the Scorecards . . . . .	64
C.4	Effects of Information Intervention on Bribes . . . . .	66
C.5	Effects of Scorecard on Applicant Satisfaction . . . . .	66
<b>D</b>	<b>Additional Tables and Figures</b>	<b>67</b>
<b>II</b>	<b>The Effects of Social Movements: Evidence from #MeToo (Joint with Ro'ee Levy)</b>	<b>96</b>
<b>1</b>	<b>Introduction</b>	<b>96</b>

<b>2</b>	<b>Underreporting of Sexual Misconduct and the MeToo Movement</b>	<b>100</b>
2.1	Reporting of Sexual Misconduct . . . . .	100
2.2	The MeToo Movement . . . . .	101
<b>3</b>	<b>Identifying the Effect of the Movement: Analysis of International Data</b>	<b>103</b>
3.1	Data . . . . .	103
3.2	Empirical Strategy . . . . .	104
3.3	Results . . . . .	106
3.4	Persistence of the Effect over Time . . . . .	108
3.5	Placebo Tests . . . . .	109
3.6	Robustness Checks . . . . .	109
<b>4</b>	<b>Heterogeneity and Effect on Arrests: Analysis of US data</b>	<b>111</b>
4.1	Data . . . . .	111
4.2	Empirical Strategy . . . . .	112
4.3	Results . . . . .	114
4.4	Robustness: Matrix Completion Method . . . . .	119
<b>5</b>	<b>Mechanisms and Interpretation</b>	<b>120</b>
5.1	Changes in the Incidence of Sexual Crimes . . . . .	120
5.2	Changes to Laws and Government Policy . . . . .	121
5.3	Changes in Awareness and Beliefs . . . . .	122
<b>6</b>	<b>Conclusions</b>	<b>123</b>
	<b>Appendices</b>	<b>141</b>
<b>A</b>	<b>Data Processing</b>	<b>141</b>
A.1	Crime Classification . . . . .	141
A.2	OECD Crime Data Collection and Processing . . . . .	141
A.3	Google Search Data Processing . . . . .	143
A.4	Fraction of English Speakers Data Processing . . . . .	144
A.5	NIBRS Crime Data Processing . . . . .	144



A.6 City Crime Data Processing . . . . .	145
<b>B Additional Analysis</b>	<b>146</b>
B.1 Allowing for Different MeToo Start Dates in Each Country . . . . .	146
B.2 Neighborhood-Level Heterogeneity . . . . .	147
<b>C Additional Figures and Tables</b>	<b>148</b>

## List of Figures

I.1	Value of Land, Record of Rights, Faster Processing, and Bribe Payments . . . . .	38
I.2	Overview of randomization and data collection . . . . .	39
I.3	Fraction of Applications Processed within 45 Working Days . . . . .	40
I.4	Effect of Scorecards by Time Since the Start of Experiment . . . . .	41
I.5	Histogram of Processing Times by Treatment . . . . .	42
I.A.1	Digital Government Capacity by Income Group . . . . .	68
I.A.2	Application Process for Successful Application . . . . .	69
I.A.3	Stated Reasons for Bribe Payments . . . . .	70
I.A.4	Example of Performance Scorecard . . . . .	71
I.A.5	Example Peer Performance List . . . . .	72
I.A.6	Information Pamphlet Given in Information Intervention . . . . .	73
I.A.7	Fraction of Applications Processed within 45 Working Days . . . . .	74
I.A.8	Comparison of Estimated Bribes to Transparency International Bangladesh Survey .	75
II.1	Share of Assaults Reported to the Police . . . . .	125
II.2	Google Search Interest in the OECD . . . . .	126
II.3	Immediate Search Interest in the MeToo Movement . . . . .	127
II.4	Crimes Reported over Time . . . . .	128
II.5	Placebo Tests, Setting the Start Date of the MeToo Movement in Every Second Quarter from Q2 2010 to Q4 2017 . . . . .	129
II.A.1	Newspaper Coverage . . . . .	149
II.A.2	Relationship Between Google Search Interest and Knowledge about the MeToo Move- ment . . . . .	150
II.A.3	Search Interest by the Strength of the MeToo Movement . . . . .	151
II.A.4	Variation in MeToo Interest Across US States . . . . .	152
II.A.5	Matrix Completion Results . . . . .	153

## List of Tables

I.1	Summary Statistics . . . . .	43
I.2	Effect of Scorecards on Processing Times . . . . .	44
I.3	Effect of Peer Performance List . . . . .	45
I.4	Effect on Visits and Time Spent by Applicants . . . . .	46
I.5	Effect on Bribe Payments for Application Processing . . . . .	47
I.6	Effects on Processing Times, Visits, and Time Spent by Office Baseline Performance . . . . .	48
I.7	Effect on Bribes by Office Baseline Performance . . . . .	49
I.A.1	Balance of Randomization: Administrative Data . . . . .	76
I.A.2	Balance of Randomization: Survey Data . . . . .	77
I.A.3	Testing Prediction from Monopolistic Price Discrimination Model . . . . .	78
I.A.4	Effect of Information Treatment and Scorecards on Bribes . . . . .	79
I.A.5	Robustness to Imputation Technique: Effect on Processing Times . . . . .	80
I.A.6	Robustness to Functional Form Assumption: Effect on Processing Times . . . . .	81
I.A.7	Robustness to Alternative Specifications: Effect on Processing Times . . . . .	82
I.A.8	Effect on Office by Month Level Outcomes . . . . .	83
I.A.9	Robustness to Alternative Specifications: Effect on Bribes . . . . .	84
I.A.10	Robustness to Measures of Baseline Performance: Effect Heterogeneity . . . . .	85
I.A.11	Effects on Expected Processing Time . . . . .	86
I.A.12	Effect on Bureaucrat Transfers . . . . .	87
I.A.13	Treatment Effects on Survey Attrition . . . . .	88
I.A.14	Lower Lee Bounds for Effects on Bribes . . . . .	89
I.A.15	Comparison of Effects in Administrative and Survey Data . . . . .	90
I.A.16	Effect by Date of E-Governance System Installation . . . . .	91
I.A.17	Robustness to Excluding Applications Potentially Affected by Applicant Survey . . . . .	92
I.A.18	Effect on Applications Received and Land Size . . . . .	93
I.A.19	Effect on Rejections . . . . .	94
I.A.20	Effect on Applicant Satisfaction . . . . .	95
II.1	Effect of the MeToo Movement During the First Six Months . . . . .	130

II.2	Persistence of the Effect in Countries with a Strong MeToo Movement . . . . .	131
II.3	Robustness Checks . . . . .	132
II.4	Effect of the MeToo Movement on Sexual Crimes in the US . . . . .	133
II.5	Effect of the MeToo Movement by Relationship and Crime Type . . . . .	134
II.6	Effect of the MeToo Movement by the Lag Between the Occurrence and Reporting Dates . . . . .	135
II.7	Effect of the MeToo Movement by Victim and Offender Demographics . . . . .	136
II.8	Effect of the MeToo Movement by County Demographics . . . . .	137
II.9	Effect of the MeToo Movement on Arrests . . . . .	138
II.10	Effect on Crimes that Occurred Before the MeToo Movement Started . . . . .	139
II.11	Change in Beliefs Regarding Sexual Harassment . . . . .	140
II.A.1	Effect of the MeToo Movement, Using Different MeToo Start Dates . . . . .	154
II.A.2	MeToo Movement Start Date by Country . . . . .	155
II.A.3	Definition of the Neighborhood Used by City . . . . .	156
II.A.4	Effect of the MeToo Movement by Neighborhood . . . . .	157
II.A.5	Data Sources for international_data . . . . .	158
II.A.6	Effect of the MeToo Movement in the US with Crime Aggregated by Offense Types .	159
II.A.7	Effect of the MeToo Movement by City . . . . .	160
II.A.8	Persistence of the Effect in the US . . . . .	161
II.A.9	Effect of the MeToo Movement on Clearance . . . . .	162
II.A.10	Effect of Crime Covariates on Changes in the Sexual Assault Arrest Rate . . . . .	163

## Acknowledgments

I am very grateful to my advisors Ahmed Mushfiq Mobarak, Rohini Pande, and Mark Rosenzweig for guiding me through the process of writing this dissertation. Mushfiq encouraged me to be ambitious and provided the support and contacts to fulfill that ambition. Rohini provided valuable feedback on both the big and small issues in my research, and reminded me about the value of my work when I did not see it myself. Mark instilled in me a scientific mindset and tirelessly insisted on, and helped me with, applying a rigorous logic to my research process. I have also greatly benefited from comments and suggestions from faculty in Yale's Development Economics and Public/Labor Economics groups. Michael Boozer, Dan Keniston, Costas Meghir, Nick Ryan, Joseph Shapiro, Ebonya Washington, and Fabrizio Zilibotti have been especially important in shaping this dissertation.

A large number of Yale graduate students also improved this dissertation by providing valuable feedback and moral support. I am particularly grateful to have had Gaurav Chiplunkar, Eduardo Fraga, Ro'ee Levy, Oren Sarig, Jeff Weaver, Jaya Wen, and Lucas Zavala as friends and colleagues. The best professional decision I made during my PhD was to co-author with Ro'ee. His wisdom, work ethic, and friendship made my time as a graduate student both more productive and happy.

Carrying out a large-scale field experiment in Bangladesh was only possible because of IPA Bangladesh's assistance. I am forever indebted to Mehrab Ali, Ashraful Haque, Alamgir Kabir, and Shahida Khala, as well as numerous field managers and enumerators, for their tremendous work and dedication enabling the experiment to succeed. The project would not have happened without the willingness to collaborate from a2i and the Ministry of Land of the Government of Bangladesh. In particular, the innovative vision of Anir Chowdhury and the continuous support from Enamul Haque and Humayun Kabir were crucial for this project. While working on this dissertation, I have also had the pleasure to work with excellent research assistants: Calvin Jahnke, Kamila Janmohamed, Mahzabin Khan, and Ashraf Mian. All of them improved the quality of the research.

The research in the dissertation received financial support from the JPAL Governance Initiative (GR-0861), the International Growth Centre (31422), the Yale Economic Growth Center, the Yale Department of Economics, the MacMillan Center for International and Area Studies, the Weiss

Family Fund, Y-RISE, and the Tobin Center for Economic Policy. The Sylff fellowship supported my doctoral studies.

I am grateful to my family, Jan, Åsa and Erling, for always being interested in my work, but also reminding me that work is not everything in life, tack! I am lucky to have amazing parents-in-law who with their kindness and boundless hospitality made Kolkata feel like a home to me, thank you Malabika mashi and Prabir uncle. Finally, I want to thank Sahana and Tara. Sahana, for her love, constant support, and for believing in me even when I don't believe in myself, and Tara for always bringing joy and meaning to my life.

## **Part I**

# **Service Delivery, Corruption, and Information Flows in Bureaucracies: Evidence from the Bangladesh Civil Service**

## **1 Introduction**

The state's capacity to implement its policies, secure property rights, and provide basic public services is paramount for economic development. To have this capacity, the state needs a functioning bureaucracy of government officials motivated to carry out their tasks. For career civil servants, compressed wage structures, secure employment, and opportunities for rent extraction through corruption often lead to weak or counterproductive incentives, especially in low- and middle-income countries. While explicit incentive structures, such as pay-for-performance contracts, can change the behavior of government officials, they are often hard to implement without unintended consequences (Finan et al., 2017). Furthermore, political constraints often prevent the introduction of explicit incentive structures altogether. However, the lack of explicit incentives does not mean that civil servants have no incentives. Supervisors in government bureaucracies often influence future postings and career paths of lower-level bureaucrats, which can be a strong motivating factor for civil servants (Khan et al., 2019). Furthermore, bureaucrats may have strong intrinsic motivations to perform their jobs well (Banuri and Keefer, 2013; Cowley and Smith, 2014).

Providing better information flows within bureaucracies about individual officials' performance may improve existing incentives by allowing supervisors to align postings and promotions more closely with job performance. Regular feedback may also increase officials' intrinsic motivation by making their own performance more salient to themselves. Furthermore, the flexible interpretation of information that is not directly tied to explicit incentives may avoid some of the common pitfalls of explicit incentives structures such as the neglect of tasks not measured by the performance indicators and opposition from individuals within the organization leading to poor implementation

(Banerjee et al., 2020). Historically, high-frequency information on bureaucrat performance has often been expensive to collect, but e-governance systems can substantially reduce this cost and increase the data quality (Singh, 2020). As low- and middle-income countries have expanded their digital capabilities, this has created new opportunities for improved information systems in the management of government officials.

This paper focuses on the processing time of applications for changes to government land records in Bangladesh. An update to the government land records has to be made every time a parcel of land changes owners and is necessary for the issuance of a land title to the new owner. Updated land records are essential for individuals to have secure property rights over land. Land disputes are one of the most severe legal problem in Bangladesh, with 29% of adults having faced a land dispute in the past four years (Hague Institute for Innovation of Law, 2018). Slow public service delivery is also a significant problem in Bangladesh. For example, only 56% of land record change applications in my control group are processed within a 45 working day time limit mandated by the government. Furthermore, faster service provision is a commonly stated reason for bribe payments, suggesting that slow service delivery on average may cause corruption as some firms and citizens pay bribes to avoid having to wait for their services.<sup>2</sup>

In an experiment with the Bangladesh Civil Service, I provide information regarding junior civil servants' performance using monthly scorecards sent to the civil servants themselves and their supervisors. The scorecards are designed to reduce delays in the processing of applications for land record changes and are based on data from an e-governance system. There are two performance indicators shown on the scorecards: the number of applications processed within the official time limit of 45 working days and the number of applications pending beyond that limit. The scorecards also show the bureaucrats' relative performance on these indicators, compared to all other bureaucrats in the experiment. The intervention is randomized at the level of the land office, and there is only one civil servant per office. The experiment was carried out at a large scale and involve 311 land offices (59% of all land offices in Bangladesh), which serve a population of approximately 95 million people.

The scorecards had a meaningful effect on bureaucrats' behavior. Using administrative data on

---

<sup>2</sup>Among households in Bangladesh reporting having paid a bribe for a public service, 23% stated that "timely service" was one of the reasons for paying the bribe (Transparency International Bangladesh, 2018).



more than a million applications, I estimate that the scorecards increase the share of applications processed within the time limit by 6 percentage points or 11%. The effect starts almost immediately after the scorecards are first sent out and is present for the 16 month period of the experiment. The scorecards also decreased the average processing time of applications by 13% and the applicants' visits to government offices by 12%. The effects are almost entirely driven by bureaucrats in offices with a below-median performance at baseline, improving their performance. This result shows that improving the information flows within a bureaucracy can change bureaucrats' behavior, even without explicit incentive structures.

Since the scorecards were sent to both the bureaucrats and their supervisors, there might be two different mechanisms for the effect on behavior. First, the bureaucrats may care about their reputation among their supervisors, potentially because of the influence the supervisors have over their careers. Second, the scorecards may also change bureaucrats' behavior by causing a sense of shame or pride through making their absolute and relative performance more salient to the bureaucrats themselves. For ease of exposition, I will refer to these two concerns as *reputational concerns*. While I cannot distinguish between these two mechanisms, I use a variation of the scorecard to test if peer effects from having the performance information shared among bureaucrats at the same level within the bureaucracy can motivate bureaucrats further than having the information shared only with the supervisors. I find no evidence of meaningful peer effects beyond the effect of the standard scorecards.

Some existing theories of corruption suggest that the average speed of public service delivery is causally and negatively related to bribes since, when average processing times are long, some applicants pay to avoid having to wait (Leff, 1964; Rose-Ackerman, 1978; Kaufmann and Wei, 1999). I conduct a survey among applicants and use the experimental variation in processing times to test theories of how they are related to corruption. Overall, the scorecards did not decrease bribe payments. The point estimate of the effect is an increase of BDT 1,046 (~USD 12) or 17%, and the lower bound of the 95% confidence interval is a decrease of 3%.<sup>3</sup> The increase comes from a positive effect on the bribe amounts reported (intensive margin) with no effect on the fraction of applicants reporting bribes (extensive margin). Using an experimental information intervention among surveyed applicants, I rule out that the lack of a decrease in bribes is due to the information

---

<sup>3</sup>Throughout the paper, I use a USD/BDT exchange rate of 84.3, the average exchange rate during the experiment.

about the improved processing times not yet having disseminated among applicants.

The positive effect of the scorecards on bribe payments is concentrated among the offices that were over-performing at baseline, i.e., the offices for which the scorecards have no effect on processing times. In the under-performing offices, where scorecards improve processing times, they do not affect bribes. This is inconsistent with a causal relationship between average processing times and bribes since processing times can improve without bribes changing, and bribes can increase without processing times changing.

I propose a model in which bureaucrats trade-off reputational concerns, bribe money, and the utility cost of effort. The bureaucrats' reputation is determined by their visible job performance along two dimensions, delays and bribe extraction, which are only imperfectly observable to supervisors. The scorecards increase the visibility of delays and thereby make them more important for reputation. For under-performing bureaucrats, this also means that the scorecards decrease their reputation. Therefore, the model predicts that under-performing bureaucrats reduce delays by providing more effort. The model also predicts that when the scorecards highlight the already good performance of over-performing bureaucrats, this relaxes their reputation constraint, allowing them to increase bribes. Furthermore, the model is consistent with the result that the scorecards do not affect delays for over-performing bureaucrats. For them, more visible delays increase incentives to avoid delays (substitution effect), but the increase in reputation makes the marginal importance of reputation smaller (income effect), so the overall effect is ambiguous.

This paper contributes to four strands of literature. First, it contributes to the literature on how incentives shape bureaucratic performance. There is an extensive literature on both monetary and non-monetary explicit incentives (e.g., Ashraf et al., 2014; Khan et al., 2016, 2019). This paper contributes to the growing literature on the effects of information flows within government bureaucracies (Dodge et al., 2018; Muralidharan et al., 2020; Dal Bó et al., 2019; Callen et al., 2020; Banerjee et al., 2020). In particular, I show that information flows about individual civil servants' performance can improve public service delivery even without explicit incentives and that this effect is persistent over time. This could be due to long-term career concerns of bureaucrats (Niehaus and Sukhtankar, 2013a; Bertrand et al., 2020) or a sense of shame or pride internal to the bureaucrats themselves (Allcott, 2011; Dustan et al., 2018). However, I find no evidence that reputational concerns *among* bureaucrats at the same level in the organizational hierarchy are a

substantial motivating factor (Mas and Moretti, 2009; Cornelissen et al., 2017). My model suggests that reputational concerns provide incentives for performance and limit the amounts of bribes collected by bureaucrats. However, the model also shows how improving the relative reputation of individual bureaucrats can lead them to perform worse and be more corrupt.

Second, the paper provides empirical evidence on the connection between corruption and the speed of public service delivery, or more generally, red tape. Slow service delivery is positively associated with corruption (Kaufmann and Wei, 1999; Freund et al., 2016), and applicants may have to pay bribes to increase processing speed for services (Bertrand et al., 2007). In the mainly theoretical literature on why the speed of service delivery and corruption are associated, different models lead to drastically different policy conclusions. One view is that corruption allows firms and individuals to circumvent excessively onerous bureaucratic hurdles (Leff, 1964; Huntington, 1968). In this view, rooting out corruption would decrease the speed of service delivery and increase inefficiencies of excessive bureaucratic control. An opposing view is that corruption is the driver of red tape and delays in public services, as making the de-jure regulation more onerous allows government officials to extract more bribes (Myrdal, 1968; Rose-Ackerman, 1978; Kaufmann and Wei, 1999).<sup>4</sup> According to this view, we could improve service delivery by eliminating corruption. According to both views, we could reduce corruption by providing services with fewer delays to everyone. I contribute to the literature by showing that, in this context, increasing the average speed of service delivery does not decrease bribe payments and that there is no evidence of a causal relationship between the average speed of service delivery and bribes.

Third, the paper contributes to the literature on the determinants of bribe amounts. In some settings, bribe payers' outside option and ability to pay constrain bribe amounts (Svensson, 2003; Bai et al., 2019), potentially leaving little room for applicant complaints or monitoring to reduce corruption (Niehaus and Sukhtankar, 2013b). In other settings, monitoring has been effective in reducing corruption (Reinikka and Svensson, 2005; Olken, 2007). I show that, in this context, individual bureaucrats *can* increase bribes and that bribes do not just reflect the difference between the official fee and the applicants' willingness or ability to pay for the service. Instead, my model highlights how bureaucrats' concerns about reputation or shame constrain bribes, explaining why

---

<sup>4</sup>In Banerjee (1997) and Guriev (2004), both corruption and red tape emerge from the nature of public service provision due to the principal-agent problem between the government and its bureaucrats. The experimental results can neither reject nor confirm these models.

bribes are substantially below applicants' willingness to pay for the service. The results are also consistent with the literature on moral licensing showing that when past pro-social behavior is made more salient, individuals tend to act less altruistically (Sachdeva et al., 2009; Clot et al., 2018).

Finally, the paper is related to the literature on the effects of e-governance in settings with low government capacity. In some cases, e-governance systems have improved government efficiency and reduced corruption (Banerjee et al., 2020; Lewis-Faupel et al., 2016). While in others, they have not had substantial benefits and wasted scarce government resources (World Development Report, 2016). This paper does not evaluate an e-governance system as a whole. Instead, it provides evidence on the untapped potential in the data that e-governance systems generate for the management of government officials.

The rest of the paper is organized as follows. Section 2 describes the context, experimental interventions, and data. Section 3 describes the empirical strategy used to analyze the experiment. Section 4 presents the estimated effects of the scorecards on processing times and bribes. Section 5 discusses how the results relates to existing theories of the relationship between the speed of service delivery and corruption. Section 6 proposes a model of bureaucratic behavior explaining the results. Section 7 concludes by discussing policy implications.

## **2 Context, Experimental Intervention, and Data**

The context of this study is land record changes in Bangladesh, and specifically the time it takes to process applications for such changes. Maintaining an updated record of land ownership is crucial for secure property rights, and globally it is an example of a public service that is almost exclusively provided by the state. More generally, the timely provision of public services is an important aspect of government capacity. The speed of public service provision is a key determinant of a country's score in the World Bank's annual *Doing Business* report. Timely public service provision and policy implementation has been shown to be positively associated with poverty reduction (Djankov et al., 2018), trade (Djankov et al., 2010), entrepreneurship (Klapper et al., 2006), and economic output (Nicoletti and Scarpetta, 2003; Djankov et al., 2006). Several countries, such as India and Russia, have explicitly stated goals to reach a certain *Doing Business* ranking, showing the importance that governments in low- and middle-income countries place on increasing the speed of public service

delivery.<sup>5</sup>

The scorecard intervention is made possible by a recently implemented e-governance system that bureaucrats in Bangladesh use to process applications for land record changes. Appendix Figure I.A.1 shows how governments in low- and middle-income countries have expanded their digital capacity compared to high-income countries in four important areas of governance. The figure shows that that public services provided using e-governance systems are now commonplace, if not the norm, even outside high-income countries.

## 2.1 Land Record Changes in Bangladesh

When a parcel of land changes owners in Bangladesh, either through sale or inheritance, the official land record has to be changed and a new record of rights issued to the new owner. Land record changes (called "mutations" in Bangladesh) are conducted by civil servants holding the position of Assistant Commissioner Land (ACL). Throughout the paper, I am referring to ACLs as the *bureaucrats*. ACL is a junior position in the Bangladesh Administrative Service, the elite cadre of the Bangladesh Civil Service. Each ACL heads a sub-district (*Upazila*) land office. The ACL is directly supervised by an Upazila Nirbahi Officer (UNO), the most senior civil servant at the sub-district level. The UNO is then supervised by a Deputy Commissioner (DC), the most senior bureaucrat at the district level. The UNO has substantial power over the ACL's future career through an Annual Confidential Report regarding the performance of the ACL that the UNO submits to the Ministry of Public Administration. Throughout the paper, I am referring to the UNOs and DCs as the *supervisors*.

A bureaucrat typically holds the position of ACL for one to two years and when an ACL is transferred, it is often to a different position within the bureaucracy. For example, of the 615 ACLs I observe in my administrative data, only 10% held the position of ACL in more than one land office.

The de-jure process for making a land record change is visually represented in Appendix Figure I.A.2. To make a land record change, the new owner must apply for such a change at the sub-district land office where the land is located. Hence, there is no competition between land

---

<sup>5</sup>India aims to be in the top 50 (<https://www.livemint.com/Politics/D8U9SSxwJ741OH7CxEYZeO/India-unlikely-to-see-significant-rise-in-Doing-Business-ran.html>) while Russia aims to be in the top 20 (<https://russiabusinesstoday.com/economy/russia-advances-in-doing-business-ranking-but-fails-to-enter-top-20/>).

offices for applicants. The application is then inspected by the office staff, who verify that the application has the required documents. The application is then sent to the local (*Union Parishad*) land office of the area where the land is located. The local land office is the lowest tier of land offices and is staffed by a Land Office Assistant who verifies the applicant's claim to the land by meeting with the applicant and visually inspecting the land. The Land Office Assistant then writes a recommendation on whether to accept or reject the application to the sub-district land office. The application is then verified against the existing government land record. Finally, a meeting is held between the ACL and the applicant where, the application is formally approved. The applicant then pays the official fee of BDT 1,150 (USD ~14) for the issuance of the new record of rights. When the applicant has paid the fee, the new record of rights is issued and given to the applicant. The sub-district land office also changes the official government land record to reflect the new ownership. The Government has mandated that land record changes should take no more than 45 working days, but in practice delays beyond this time limit are common. In my data, only 56% of applications in the control group were processed within the time limit and the average processing time among processed applications was 52 working days.

### **2.1.1 Bureaucrats' Discretionary Powers and Corruption**

In practice, it is common for applicants to also pay bribes beyond the official fee to get their application processed. Figure I.1 shows that among the applicants in my survey, the average estimated bribe for a typical applicant was BDT 6,731 (~USD 80).<sup>6</sup> Appendix Figure I.A.3 shows that when asked, the most common response to the question of why a bribe was paid is akin to "to get the work done" (39%), the second most common response is akin to "to avoid hassle" (39%), and the third most common is akin to "for faster processing" (10%). This highlights that the bureaucrat has decision making power over the application along two dimensions. First, they can decide whether to accept or reject the application. Second, they can take actions to speed up or slow down the application as well as create more or less hassle for the applicant. Figure I.1 shows that the average stated valuation of getting the record of rights is BDT 1,594,664 (~USD 18,917), almost as high as the average estimated market value of the land itself. These valuations are more than two

---

<sup>6</sup>Appendix Figure I.A.8 shows that this estimate is similar to an estimate by Transparency International Bangladesh of the average bribe paid for a land record change.

orders of magnitude larger than even the highest estimate of the average bribe payments.

On average, the applicants in my survey state that their willingness to pay for having their application processed within seven days (the shortest reasonable processing time) is BDT 2,207 (~USD 26). Since this number is substantially lower than the average bribe paid by those reporting a non-zero bribe and the average estimated typical bribe, it is clear that applicants are not just paying for faster processing. Most likely they are also paying for getting the approval.

## **2.2 E-governance System for Land Record Changes**

In February 2017, a new e-governance system for land record changes was introduced, with the goal of simplifying the process of land record changes for both the applicants and the civil servants processing the applications. The system was gradually implemented in sub-district and local land offices. As the e-governance system had recently been implemented at the time of the experiment, not all applications were processed using the e-governance system even in the sub-district offices where it had been installed. The main reason for this was that not all local land offices within the sub-district had had the e-governance system installed.<sup>7</sup>

The e-governance system generates administrative data on each application made in the system. Specifically, this administrative data can be used to assess the adherence to the rule that all applications should be processed within 45 working days. However, until the start of the experiment, this data was not presented in a format enabling evaluation of the degree of adherence to this rule or the performance of specific sub-district land offices or ACLs.

## **2.3 Experimental Intervention: Performance Scorecards**

Together with the Government of Bangladesh, I designed a monthly performance scorecard addressed to the ACL and sent to randomly selected sub-district land offices, as well as to the offices of the UNO and the DC, the ACL's two direct superiors. The scorecard is intended to decrease delays in application processing for land record changes. Appendix Figure I.A.4 presents an example of a performance scorecard.

The scorecard evaluates the ACL's performance using two performance indicators. The first

---

<sup>7</sup>Other reasons cited for using the paper-based system were problems with internet connectivity, new officials not yet trained in using the e-governance system, and temporary problems with the e-governance server.

indicator is the number of applications disposed within 45 working days in the past month, where a higher number indicates a better performance. The second indicator is the number of applications pending beyond 45 working days at the end of the month, where a lower number indicates a better performance. The scorecard shows both these numbers as well as the average numbers for all sub-district land offices in the experiment. The scorecard also provides the office's percentile ranking for each indicator, with a short sentence reflecting the performance. Finally, to make the score easily understandable and more salient, a thumbs-up symbol is put next to percentile rankings between the 60th and the 100th percentile, while a thumbs-down symbol is put next to percentile rankings from the 0th percentile to the 40th percentile. Two versions of the scorecard, one in English and one in Bengali, were sent out in the first two weeks of each month with information based on the previous calendar month's e-governance data. Offices in the treatment group were not informed that they would receive a scorecard before the start of the treatment, but the first scorecard was followed by a phone call to the ACL where the indicators were explained and the ACLs could ask questions about the scorecard. The scorecards are also accompanied by an explanatory note showing how the numbers in the scorecard are calculated and a phone number to call to ask questions about the scorecard.

### **2.3.1 Additional Intervention: Peer Performance List**

To test for peer effects, an addition was made to the scorecards for 77 randomly selected treatment offices in September 2019, a year after the first scorecards were sent out. The purpose was to test if there was an additional effect, beyond the effect of the scorecard, stemming from a bureaucrat's performance being observable to the bureaucrat's peers at the same position in the organizational hierarchy. For a randomly selected group of 77 offices within the offices already receiving the performance scorecards, a list of the percentile rankings of the two performance indicators for all 77 offices was added to the scorecard. Appendix Figure I.A.5 shows an example of the first page of such a list. The main difference between receiving the typical scorecard and the scorecard with the list of performances was that for the offices that received the list of performances, their performance was observable not just to them and their supervisors but also to 76 of their fellow ACLs.



## 2.4 Randomization

Figure I.2 provides a visual overview of the randomized interventions and data sources. The randomization was done in two waves. In August 2018, 112 land offices were using the e-governance system. In the first randomization wave, 56 of these offices were randomly chosen to receive the performance scorecards, while 56 were assigned to the control group.<sup>8</sup> In April 2019, 199 additional offices had started to use the e-governance system and a second randomization wave was carried out to increase the experiment's sample size. The second randomization wave extended the treatment to 99 new offices while 100 new offices were added to the control group.<sup>9</sup> The additional list of peer performances was added to the scorecard for 77 randomly selected offices receiving the scorecards in September of 2019. The scorecards were sent out until March 2020, when the outbreak of COVID-19 caused an end to the scorecards being sent out.

Both randomization waves were stratified by the number of applications processed within 45 working days in the two months preceding the randomization and the number of applications pending for more than 45 working days at the end of the month preceding the randomization. For the first randomization, another binary variable for being a land office where the e-governance system was fully implemented, meaning that no applications were conducted using the traditional paper-based method, was also used for stratification. In the second randomization, the total number of applications received since the installation of the e-governance system was used as a stratification variable. The randomization of offices into receiving the peer performance list was done among the 155 offices receiving the scorecards using the same stratification variables as in the second randomization wave. For more information about the randomizations see Appendix Section B.1.

## 2.5 Data

I use two main data sources, administrative data from the e-governance system and data from a survey conducted among applicants in the 112 land offices that were part of the first randomization wave. I use the administrative data to generate the performance scorecards as well as evaluating the effects of the scorecards. Table I.1 shows summary statistics for both data sets. This table contains all observations from both treatment and control offices that are used in the analysis. For a

---

<sup>8</sup>The first randomization was carried out by the author on 14 August 2018.

<sup>9</sup>The second randomization was carried out by the author on 10 April 2019.

discussion of the balance of randomization, see Section 2.6.

### 2.5.1 Administrative Data

The observations in the administrative data are at the application level. The data contains information about in which land office the application was made, the application start date, the date it was processed as well as the decision to accept or reject the application. The administrative data also contains information on the size land plot for which the change is being made.<sup>10</sup> The administrative data was downloaded from the e-governance system at the beginning of each month from August 2018 until October 2020.

For the main analysis, I use administrative data for applications made from 13 August 2018, one month before the start of the experiment, until 20 January 2020. From 26 March 2020 and onwards the COVID-19 outbreak in Bangladesh substantially increased processing times for land records changes as measured by calendar days but also resulted in a large number of general holidays, increasing the difference between calendar days and working days. At this time, the scorecard intervention was also stopped. Therefore, I do not include applications made after 20 January 2020, 45 working days before the start of the general holiday caused by COVID-19. Ending the data at this point precludes the holiday from affecting one of the main outcomes, if the application was processed within the 45 working day time limit or not.

I impute the processing times for the 6% of applications that have not yet been processed. The imputed value is the mean of actual processing times that are larger than the number of working days the application that I am imputing the processing time for has been pending.<sup>11</sup> The data set in the main analysis contains 1,050,924 applications from all 311 offices. Appendix Section B.2.1 provides more information about the administrative data.

---

<sup>10</sup>The full administrative data set also contains more information about the applicants, but this data is not available for research purposes due to privacy concerns.

<sup>11</sup>This procedure is conservative in two ways. First, it reduces any effect on processing times generated by the scorecards since the same mean is used to impute values in both the treatment and control areas. Second, the mean used to impute processing times in this procedure likely underestimate the time it will take to process these applications on average since it is the mean of applications that *have already been processed*, which is likely to be less than the actual average time it will take to process all applications including those currently pending. Since the point estimate of the scorecards' effect on the share of applications being pending is a decrease of 0.9 percentage points, using these imputed values creates a conservative estimate of the effect of the scorecards on processing times.

### 2.5.2 Survey Data

The survey data was collected in two rounds from applicants who applied in the 112 offices that were part of the first wave of randomization. The sample of applicants was created by placing surveyors outside land offices and interviewing all applicants entering the office for the purpose of a land record change application, regardless of what stage in the application process they were at. The surveyors stayed outside a specific office for at least two days and until they had completed at least 20 interviews. The follow-up interview was conducted by phone approximately three months after the initial interview. Surveyors were not informed about which offices had received the scorecards or if they were calling a respondent from a treatment or control office.

Out of 3,696 people approached, a total of 3,370 applicants were successfully interviewed in the first round interview outside of the land offices. Out of those interviewees, 3,018 were successfully re-interviewed in the follow-up phone interview, resulting in a total attrition rate of 18%. The estimated effect of the scorecards on the attrition rate was 3 percentage points and marginally statistically significant at the 10% level. However, in Appendix Section B.2.3 I show that this differential attrition is not sufficiently large to substantially affect the main findings from the survey data. More information about the survey data can be found in Appendix Section B.2.2.

The initial interview focused on the details of the application, the applicant's expectation for the application processing time, the applicant's willingness to pay for faster processing, as well as basic information about the applicant. The follow-up interview focused on the outcome of the application and the payments, above the official fee, that the applicant had made in relation to the application.

Data on bribe payments was collected using two different questions. The first question asked what the typical bribe payment is "for a normal person like yourself." If the respondent were willing to answer this question, the amount, whether zero or positive, was recorded as the variable *typical payment*. 63% of respondents provided an answer to this question and the average response was BTB 6,731 (~USD 80) or 1.5 months of the sample's average per capita household expenditure.<sup>12</sup> 73% of the responses were non-zero amounts.

---

<sup>12</sup>All continuous variables from the survey are winsorized at the 99th percentile. Averages are calculated using observations weighted by the inverse of the number of observations in each office, making the estimates representative for the average land office.

The second set of questions asked about each actual payment made by the applicant to any government official or agent assisting with the application. The outcome variable *reported payment* is the sum of the bribe amounts reported in each of these questions. This variable takes the value zero when no payments were reported. The most common response for respondents who were not willing to talk about payments that they had made was to report no payment. Therefore, the average reported payment is likely an underestimate of the average payment actually made. The average reported payment was BDT 1,456 (~USD 17) and 27% of respondents provided a non-zero value. Among those reporting a non-zero amount the average amount was BDT 5,283 (~USD 63).

## 2.6 Balance of Randomization

Appendix Table I.A.1 shows a balance of randomization test for the two main outcome variables from the administrative data, the fraction of applications processed within 45 working days and the average processing time. The data used is restricted to applications made at least 45 working days before the start of the experiment. Applications that were not processed by the start of the experiment were assigned an imputed processing time, using the imputation procedure described in Section 2.5.1. There are no statistically significant differences between scorecard and control offices before the start of the experiment. This is expected given that the random treatment assignment.<sup>13</sup>

Appendix Table I.A.2 shows that the scorecards did not affect the composition of applicants or applications in the survey data. This is not a traditional balance of randomization table, since the treatment may have affected which applicants decided to apply and what type of applications to make. However, I do not find any evidence for such changes in behavior. I find no statistically significant difference in the age or income of the applicants, or in the size or value of the land that the applications are for. Furthermore, there are no substantial differences between the stages that the applications are in at the time of the first interview. When using the regression specification from Equation 1 on this data, the effect of the scorecards is not significant at the 5% for any of the outcome variables, and significant at the 10% level only for land value.<sup>14</sup>

<sup>13</sup>Using the empirical strategy described in Section 3.1 on the data from before the start of the experiment also generates statistically insignificant estimates of the effect of the treatment on the outcome variables. Furthermore, an F-test of joint significance for the explanatory power of the outcome variables on the treatment variable cannot reject the null of no explanatory power (p-value: 0.69).

<sup>14</sup>F-tests of joint significance for the explanatory power of the outcome variables on the treatment variable cannot reject the null of no explanatory power (p-value: 0.73).

## 2.7 Additional Intervention: Providing Information to Applicants

Together with the in-person survey, an intervention providing additional information to applicants was also carried out on randomly selected days in each office where the survey took place. The motivation behind this intervention was to ensure applicants knew about the improvements in processing times. While it is likely that this information would eventually have spread, in the short-term, information about changes to bureaucrat behavior may not yet have disseminated. If the applicants are not aware of the improvements in processing times, the long-term effects on bribe payments may not yet have been realized. To speed-up the dissemination process, and potentially reach the long-term effect of the scorecards faster, the surveyors randomly provided information about increased processing speeds on half of the days that the in-person survey was conducted. The surveyors used an information pamphlet to inform applicants that the median processing time for all land offices had been substantially reduced over the past six months and that a new e-governance system had been installed. The information the surveyors provided was the same in both treatment and control offices. The scorecard intervention was not mentioned to applicants. Appendix Figure I.A.6 shows an English translation of the information pamphlet. I analyze the effects of this intervention when testing the predictions of models connecting processing times and corruption in Section 5.

## 3 Empirical Strategy

### 3.1 Empirical Strategy: Overall Effects

To estimate the effects of the scorecards, I use the following regression specification:

$$Outcome_{ait} = \alpha + \beta Treatment_i + Strata_i + Month_t + \varepsilon_{ait} \quad (1)$$

Where  $Outcome_{ait}$  is an outcome for application  $a$ , in land office  $i$ , made in calendar month  $t$ .  $Strata_i$  are randomization strata fixed effects. Since no randomization strata overlap the two randomization waves, these fixed effects also control for randomization wave fixed effects.  $Month_t$  are fixed effects for the month the application was made. In the survey data, all continuous variables are winsorized

at the 99th percentile.<sup>15</sup> Standard errors are clustered at the land office level resulting in 311 clusters in the administrative data and 112 clusters in the survey data. Each observation is weighted by the inverse of the number of observations in each land office. Therefore, the estimated effect is the average effect of the scorecard on a land office, the level at which the treatment was assigned. The weighting also improves the estimates' precision by making each cluster have equal weight in the analysis.<sup>16</sup>

### 3.2 Empirical Strategy: Heterogeneous Effects

To better understand the mechanisms behind the overall effects, I separate offices by their baseline performance and estimate the effect of the scorecards separately for offices performing above and below the median at baseline.<sup>17</sup> I calculate each office's baseline performance based on the average of the two percentile rankings at the time of the first scorecard. One ranking is based on the number of applications disposed within 45 working days, while the other is based on the number of applications pending for more than 45 working days. For offices in the treatment group, these are the actual rankings shown on the first scorecard, while for the control group, the rankings were not shown to the bureaucrats. I then separate all offices into *over-performers*, that were above the median average ranking at baseline, and *under-performers*, that were below the median average ranking at baseline.<sup>18</sup> Since the classification of offices only uses data from before the first scorecard was delivered, it is not affected by the treatment.

I use the following regression specification to estimate the effect of the scorecards on the two

---

<sup>15</sup>In the survey data, the application month variable is winsorized at November 2018, so that all application dates before November 2018 take the value of November 2018. A separate dummy variable controls for missing start date values.

<sup>16</sup>For a discussion of why weighting observations by the inverse of the number of observations in a cluster improves precision see: <https://blogs.worldbank.org/impactevaluations/different-sized-baskets-fruit-how-unequally-sized-clusters-can-lead-your-power>

<sup>17</sup>Heterogeneity in the effects of performance information provision between high and low performers has been recorded in several settings (e.g., Allcott, 2011; Dodge et al., 2018; Ashraf, 2019; Barrera-Orsorio et al., 2020). This was the only heterogeneity test based on office characteristics specified in the pre-analysis plan. The two other pre-specified tests for heterogeneity were based on the date of application and the application processing time. The estimates of heterogeneity in the effects along those dimensions are shown in Figure I.4 and Appendix Table I.A.3, respectively.

<sup>18</sup>I classify offices in the first randomization wave into over- and under-performers by comparing them to the median performance among these 112 offices at the time of their first scorecard (September 2018). For the offices in the second randomization wave, I compare them to the median performance of all 311 offices in the experiment at the time of their first scorecard (April 2019). This ensures that the over- and under-performer classification corresponds to if the content in the first scorecards was above or below the median of comparison groups at the time.

types of offices separately:

$$y_{ait} = \alpha + \beta_1 Treatment_i \times Overperform_i + \beta_2 Treatment_i \times Underperform_i + \gamma Overperform_i + Stratum_i + Month_t + \varepsilon_{ait} \quad (2)$$

Where  $\beta_1$  is the estimated effect of the scorecards for offices over-performing at baseline,  $\beta_2$  is the effect for offices under-performing at baseline, and  $\gamma$  is the difference between over-performing and under-performing offices in the control group.<sup>19</sup> As in the estimation of the overall effects, standard errors are clustered at the land office level and the regressions are weighted by the inverse of the number of observations in land office  $i$ .

### 3.3 Analysis of Additional Experiments and Potential Interactions

The two additional randomized interventions, the addition of peer performance lists and the information intervention to applicants, are not included in the main specification as these interventions are not the main treatments being evaluated. For the two main outcomes, delays and bribe payments, the full specifications, including the scorecard treatment, the additional randomization, and the interaction, can be found in Tables I.3 and I.A.4. These tables show that neither of the two additional experiments have substantial interactions with the scorecard treatments, validating the approach to analyze the scorecard treatment separately as outlined in Equations 1 and 2.

## 4 Results: Effects on Processing Times and Bribes

This Section shows the estimates of the effects of the scorecards on processing times, visits to land offices made by applicants, and bribes. Appendix Section C.3 investigates potential unintended consequences of the scorecards on bureaucrats' behavior and does not find evidence for any large unintended consequences.

---

<sup>19</sup>To test the hypothesis that the treatment had the same effect on offices over-performing and under-performing at baseline, I use a similar regression but where the first treatment variable is not interacted with the dummy variable for if the office over-performed at baseline. I then test the hypothesis that the coefficient on the treatment variable interacted with the dummy variable for if the office was under-performing at baseline is zero. This test's p-value is reported as "P-value sub-group diff." in the regression tables reporting the heterogeneous effects.

## 4.1 Effect on Processing Times

Table I.2 shows that the scorecards increased the applications processed within the government time limit and improved processing times overall. Each column presents the result of a regression using the specification in Equation 1. Column (1) shows the estimated effect of the scorecards on a binary variable indicating if the application was processed within the 45 working day time limit or not. The scorecards increased the fraction of applications processed within the 45 working day limit by 6 percentage points or, equivalently, 11%. Column (2) shows the estimated effect on the Inverse Hyperbolic Sine (IHS) transformation of the number of working days it took to process the application.<sup>20</sup> Column (2) estimates that the scorecards reduced the processing time by 13%.<sup>21</sup> In the data, 6% of the applications are not yet processed, and for the analysis in Column (2) I have assigned imputed processing times for these applications, using the imputation procedure described in Section 2.5.1. Appendix Table I.A.5 shows that the results are robust to different imputation techniques. In Appendix Table I.A.6 I test the robustness of the result to using different functional form assumptions for the relationship between the scorecards and processing times.

For Column (3), I create an Inverse Covariance Weighted (ICW) index of the two outcomes used in Columns (1) and (2).<sup>22</sup> The estimated effect of the scorecards on the ICW index is 0.13 standard deviations and statistically significant. In Appendix Table I.A.7 I test the robustness of this result with various alternative specifications. All alternative specification estimates are of the same sign and similar magnitude as the main estimate, but some of them are not statistically significant. Appendix Table I.A.8 shows the effects, estimated at the office by month level, on the number of applications processed within 45 working days, the number of applications pending beyond 45 working days as well as those figures corresponding percentile rankings. The point estimates

---

<sup>20</sup>The IHS transformation is used instead of the natural logarithm since 0.3% of the applications were processed on the same day as they were made and therefore have a processing time of zero working days. The results are virtually identical when dropping the applications taking zero days to process and using the natural logarithm transformation.

<sup>21</sup>The exact effect is 13 IHS points, which are approximately equivalent to log points. A 13 log point decrease is equivalent to a 12% decrease, but for simplicity, I will describe IHS points changes as percentage changes throughout the paper. Appendix Table I.A.6 shows that the result is similar when dropping the observations with processing times of zero working days and using the natural logarithm transformation.

<sup>22</sup>The ICW matrix follows the algorithm suggested by Anderson (2008) and is designed to summarize several outcome variables into one index that, for the control group, has a mean of zero and a standard deviation of one. Since there are only two outcome variables in Table I.2, the ICW index is equivalent to summing the standard deviations away from the control group mean of the two variables and rescaling the index to have a standard deviation of one in the control group. However, in tables with more than two outcome variables, the components are weighted differently to maximize information captured by the ICW index.



suggest that the scorecards improved all four of these outcome variables but the effects are not statistically significant.

#### **4.1.1 Effect Over Time**

Figures I.3 and I.4 show that there is no pattern of the effect declining over time, although the size of the effect varies between different time periods. Figure I.3 shows the fraction of applications processed within the 45 working day limit over time for the treatment and control group separately. The first dashed vertical line indicates the date 45 working days before first scorecards. The second dashed vertical line indicates the date of the first scorecards. Applications made between the first and second vertical lines may have been affected by the scorecards if they were not processed before the first scorecard was sent out. Starting for applications made a few days before the first scorecards, we see a divergence between the treatment and control group. The treatment group increased the fraction of applications that were processed within the 45 working days time limit, relative to the control group. With a few short exceptions, the treatment offices continue to have a higher fraction of applications processed within the time limit relative to the control offices until the end of the experiment. The data for the offices in the second randomization wave ends earlier relative to the start of the experiment. The third vertical dashed line marks where the data from the second randomization wave ends. To the right of this line, the graph only contains data from the offices in the first randomization wave. Appendix Figure I.A.7 shows the time lines for the two randomization waves separately.

Figure I.4 shows the results of applying the regression specification from Equation 1 to applications made in the first, second, and last third of the experiment period. The outcome variable is the ICW Index from Column (3) of Table I.2. When I split up the sample, the estimates lose some precision, but it is clear from the graph that there is no pattern of a continuous decline of the effect over time.

#### **4.1.2 Effect on the Distribution of Processing Times**

Figure I.5 shows two overlaid histograms, one for the distribution of processing times in the treatment group and one for the distribution in the control group. The figure only includes

applications that have already been processed and processing times are top coded at 200 working days. In the treatment offices, more applications were processed within the 45 working day time limit. The effect is relatively evenly spread over the whole span from 0 to 45 working days, with only a minor bunching just before the 45 working day limit. This is to be expected given that the process to approve an application is relatively long and depends on several individuals, as described in Section 2.1. This means that even if the ACL targets a 45 working day processing time, there will be a considerable spread around this target. Because of this, the ACLs may target a processing time lower than 45 working days. The figure also shows that the processing times that are reduced in frequency by the scorecards are in the whole span from 55 working days and up. This is also reasonable given that the scorecards emphasized both processing applications within the 45 working day limit and reducing the number of applications pending beyond 45 working days. Overall the spread of the effect in the distribution of processing times alleviates the concern that ACLs are "gaming" the scorecards by only speeding up the processing of applications that would otherwise have been processed within a few working days outside of the time limit.

## **4.2 Mechanisms for the Effect on Processing Times**

The scorecards increase the information the bureaucrats and the bureaucrats' supervisors have about the performance of the bureaucrat. This could improve performance through two main channels. First, the supervisors may improve the incentive structures bureaucrats are facing by facilitating better promotions and more attractive postings for those bureaucrats with good scorecards, or more generally, bureaucrats with a good overall reputation of which the scorecards are a part. This is an example of the widely studied mechanism of increased information enabling better contracts that improve output (Holmström, 1979). It is also possible that bureaucrats care about their supervisors receiving information about them for other reasons, such as the shaming effect of having a negative performance being shown to a superior.

Second, bureaucrats may change their behavior due to receiving the scorecards themselves. For bureaucrats, receiving information about their delays each month may increase this information's salience, causing it to be more important for their personal sense of shame or pride in their work.<sup>23</sup>

---

<sup>23</sup>Effects from simply being informed of one's own performance have been found for energy conservation (Allcott, 2011). On the other hand, the effects of such information provision in private organizations have been mixed, with

Since the scorecards were sent to both bureaucrats and their supervisors, I cannot separately estimate the importance of these two mechanisms and I refer to them collectively as *reputational concerns*.

In addition to the two mechanisms above, it is also possible that information flows between bureaucrats at the same level in the organizational hierarchy create an additional incentive for improved performance through peer effects (Mas and Moretti, 2009; Bandiera et al., 2010; Cornelissen et al., 2017).<sup>24</sup> I estimate the magnitude of such a peer effect, above and beyond the effect of the scorecard, by sending information about other offices' performance within a randomly selected sub-group of the offices receiving scorecards, as described in Section 2.3.1.

Table I.3 shows the effect of the peer performance list intervention on processing times. Sharing the performance information of a bureaucrat with other bureaucrats does not meaningfully improve processing times beyond the effect of the performance scorecards. Column (1) of Table I.3 shows that the estimated effect on the fraction of applications processed within the 45 working day time limit is positive but close to zero. Column (2) shows that the effect on overall processing times is negative but also close to zero.<sup>25</sup>

### 4.3 Effect on Visits to Land Office and Time Spent by Applicants

In Table I.4, I use survey data to show that the scorecard reduced the number of visits to land offices by the applicants as well as the total hours spent on making these visits. Column (1) shows that the scorecards reduced the number of visits by 1.0 visits, or 12%. Column (2) estimates that the scorecards decreased the total number of hours spent on these visits by 1.6 hours, or 7%, but this effect is not statistically significant. Column (3) estimates the effect on an ICW index of these two outcome variables showing that the effect is not statistically significant for a combination of the two variables. Appendix Section C.5 shows that the scorecards did not improve the stated satisfaction

---

several papers showing that even the direction of the effect depends on the specific circumstances (Blader et al., 2020; Ashraf, 2019).

<sup>24</sup>In addition to the context of job performance, effects of sharing information about behavior to others have shown to improve socially desirable behaviors such as voting (Gerber et al., 2008) and paying taxes (Bø et al., 2015; Perez-Truglia and Troiano, 2018).

<sup>25</sup>Columns (3) and (4) of Table I.3 use the full data set and estimate the effect of the scorecard and the peer performance list simultaneously. This is done using a dummy variable for the peer performance list treatment that takes the value of one for applications made in offices receiving the peer performance lists, made later than one calendar month before the first performance list was sent out. When estimating the effects of the scorecards without the effect of the performance list, the point estimates are similar to the effect in the main estimate but only statistically significant at the 10% level. This shows that the effect of the scorecards is not driven by the inclusion of the peer performance list.

with the application process among applicants.

#### 4.4 Effect on Bribe Payments

Table I.5 shows that the scorecards did not lead to a decrease in bribe payments. Instead, the estimated effect on bribes is positive, although this increase is not statistically significant. As described in Section 2.5.2, data on bribe payments was collected using two separate survey questions. The first question asked about the typical bribe payment "for a normal person, like yourself." When this measure is used, the column is marked as "typical." The second set of questions asked about each payment made by the applicant. When this measure is used, the column is marked "reported."

Columns (1) and (2) of Table I.5 show the effect on the amount of bribes paid. Column (1) shows that the effect on the perceived typical payment was BDT 1,046 (USD 12), a 17% increase, statistically significant at the 10% level. Column (2) estimates that the scorecards increased reported bribe payments by BDT 265, a 21% increase, but the result is not statistically significant. Columns (3) and (4) show that there is no effect on the propensity to report a non-zero bribe. This can be interpreted as the scorecards having no effect on the extensive margin of bribe payments. Another interpretation is that the intervention did not affect applicants' willingness to talk about bribe payments in the survey. In Columns (5) and (6), the sample is restricted to those who reported non-zero bribe payments. Bribe payments increased by 19% for typical payments and 23% for reported payments, with both effects being statistically significant. Again these effects have two interpretations. Either the scorecards only affected the intensive margin of bribe payments, or the scorecards increased bribe payments for at least those applicants who were willing to describe what bribes they paid but potentially also for other applicants.

The estimated effects for a range of alternative specifications for the main estimate in Columns (1) and (2) of Table I.5 are shown in Panel A of Appendix Table I.A.9. All alternative specification estimates are qualitatively similar, but some are of slightly larger magnitude and, therefore, statistically significant.

## 4.5 Heterogeneity of Results by Office Performance at Baseline

I use the empirical strategy described in Section 3.2 to understand if there are differences in the effect of the scorecard between offices over- and under-performing at baseline.

### 4.5.1 Heterogeneity in Effects on Processing Times, Visits, and Time Spent by Applicants

Table I.6 shows that the effect of the scorecard on processing times is driven by offices that were under-performing at baseline. Column (1) of Table I.6 shows that for offices that were over-performing at baseline, the estimated effect on the fraction of applications processed within the 45 working day limit is just a 0.8 percentage point increase. For offices that were under-performing at baseline, the effect is 12 percentage points, equivalent to a 30% increase. Column (2) shows that for offices over-performing at baseline, the estimated effect of the scorecard on the total processing time is a decrease of 3%. For offices that were under-performing at baseline, the effect was a decrease of 23%.

The effects in the survey data are less precisely estimated but also show that it is offices underperforming at baseline driving the effect. Column (3) of Table I.6 shows that for offices over-performing at baseline, the number of visits per applicant was reduced by 0.7 visits, while for offices under-performing at baseline, the effect was a decrease of 1.2 visits, equivalent to a 12% decrease. Column (4) shows that for offices over-performing at baseline, hours spent on the application by applicants increased by 0.4 hours while in offices under-performing at baseline the effect was a decline of 3.0 hours, equivalent to an 11% decrease.

Overall it is clear that the improvements that the scorecards led to were almost entirely driven by offices that were under-performing at baseline. Appendix Table I.A.10 shows that this result is robust to other measures of baseline performance. Panel B of Appendix Table I.A.7 shows that this result is robust to alternative regression specifications. There are several reasons for why under-performing offices may respond more to the scorecards. For example, negative performance feedback may create a stronger desire to improve one's performance for subsequent scorecards. However, it may also be the case that poorly performing offices have a larger scope for improvement since there is more "low-hanging fruit" in terms of increasing efficiency.

#### 4.5.2 Heterogeneity in the effect on bribe payments

Table I.7 shows that the positive effect on bribe payments is entirely driven by the offices that were over-performing at the start of the experiment. Column (1) shows that the effect of the scorecard on estimated typical bribe payments among offices over-performing at baseline was an increase of BDT 2,280 or 43% and statistically significant. Column (2) shows that the reported payments among offices over-performing at baseline increased by BDT 638 or 70%, and is statistically significant. The effect on offices that were under-performing at baseline is close to zero and not statistically significant. Appendix Table I.A.10 shows that this result is robust to other measures of baseline performance. Panel B of Appendix Table I.A.9 shows that this result is robust to alternative regression specifications.

This result is surprising, given that over-performing offices did not change their behavior in terms of processing times. I will discuss this result at length in Sections 5 and 6.

#### 4.6 External Validity, Potential Biases from Surveying, and Unintended Consequences<sup>26</sup>

One advantage of the design of the experiment is that it was conducted at a large scale, with more than half of Bangladesh's land offices taking part in the experiment. The large scale of the experiment makes it plausible that the results are externally valid within Bangladesh (Muralidharan and Niehaus, 2017). Offices took part in the experiment if they had the e-governance system installed. Therefore, the main concern for the external validity of the result within Bangladesh is that offices that had the e-governance system installed earlier had a larger effect than offices where the e-governance system was installed later. Appendix Section C.1 shows evidence that the effect of the scorecards on processing times was only slightly larger in the land offices that had the e-governance system installed earlier. Furthermore, using a linear prediction, the effect is predicted to be positive for all offices in Bangladesh where the e-governance system is installed.

While it is unlikely, it is possible that the survey and information intervention affected the overall effect of the scorecards. In Appendix Section C.2 I restrict the sample to applications made before the survey took place and application made in offices where there was no survey and show that there is no evidence that the survey or information intervention are drivers of the estimated

---

<sup>26</sup>Appendix Section C discusses potential biases, external validity, and unintended consequences in more detail.

effect of the scorecards on processing times.

A common problem of quantitative performance measures is that they often lead to gaming of the quantitative measures or other unintended consequences (e.g., Banerjee et al., 2008; Rasul and Rogger, 2018). In Appendix Section C.3 I test for three such potential unintended consequences that could have improved the scorecards without increasing the real service delivery speed for applicants. First, if bureaucrats allow fewer applicants to start applications, then this may improve their scorecards, provided that the lower number of applications help them process a larger share of the applications within the time limit. Second, if bureaucrats allowed applications selectively such that the average application was easier to process within the time limit, then this could have improved their scorecards. Finally, the scorecards may lead to bureaucrats making worse decisions regarding accepting or rejecting applications. Reassuringly, I do not find any evidence for any of these unintended consequences.

## **5 Implications for Theories of the Relationship between Processing Times and Bribes**

In this Section I will show how the experimental results are inconsistent with several common models of how bribes are related to delays in public service delivery, or more generally red tape.<sup>27</sup> There are several theoretical reasons for why bribes may be causally related to delays. Some of these models predict a positive causal relationship, while others predict a negative relationship. For example, fast processing times may increase the applicants' willingness to pay for the public service and enable bureaucrats to extract more bribes. Bribes may also provide a piece rate incentive for bureaucrats to process more applications and cause bureaucrats to process applications faster. Conversely, long processing time for those paying small or no bribes may enable bureaucrats to extract more bribes from applicants willing to pay to get their application processed fast. These causal relationships exist both in models where corruption is efficiency-enhancing (e.g., Leff, 1964; Huntington, 1968), as well as in models where corruption is the original cause of the slow service delivery (Myrdal, 1968; Rose-Ackerman, 1978; Kaufmann and Wei, 1999).

---

<sup>27</sup>I use the term delays, but the theories are equally applicable to other forms of red tape, such as the need for multiple visits to government offices or an excessive amount of paperwork to be filled out.

It is important to understand which, if any, of these relationships are major determinants of bribes and processing times since some of the models have opposing policy implications. If slow service delivery causes corruption, then expanding the processing capacity of the bureaucracy through more staff, better technologies, or better management, may not only improve processing times but also reduce corruption. But if bribery was to be rooted out, without addressing the underlying capacity constraints, this might lead to a worse situation for applicants if, for example, bureaucrat had less of an incentive to process their applications or people who urgently needed a service could not pay to get it faster. On the other hand, if corruption is the underlying cause for slow service delivery due to intentional delays by bureaucrats for the purpose of extracting more bribes, then providing the bureaucracy with more staff or better technology would not lead to any improvement in processing times, let alone decrease corruption. The most important policy priority should then instead be to eliminate corruption to remove the incentives for bureaucrats to intentionally delay corruption.

The model the scorecard experiment was originally designed to test was a model where the presence of corruption led to slower processing times but where faster processing times could reduce corruption. The model was similar to monopolistic price discrimination models, and in the model bureaucrats use delays strategically to maximize the total amount of bribes in the same way a monopolist would strategically decrease the quality of some goods to maximize profits (Mussa and Rosen, 1978; Maskin and Riley, 1984). The model assumes that applicants have different willingness to pay to avoid delays, but that bureaucrats cannot perfectly observe the willingness to pay of each applicant. Therefore, they intentionally delay applications from applicants only paying low bribes in order to extract more bribes from applicants with a high willingness to pay to avoid delays. The model predicts that an improvement in processing times, such as the improvement the scorecards created, should decrease bribe payments among applicants getting their applications processed the fastest. See Appendix A.2 for a more detailed description and an explicit test rejecting this model in this context.

A different type of models, are models where the government officials could extract more bribes if they wanted to, but choose not to do so because there is a trade-off between taking bribes and some other objective of the government official. This trade-off could be between taking bribes and the risk of getting caught (Becker and Stigler, 1974; Olken, 2007; Niehaus and Sukhtankar, 2013a),



but it could also be a trade-off between bribes and altruistic or social motivations for not taking bribes. In Section 6, I develop a specific such model where taking bribes hurts bureaucrats' utility through bribes negative effect on reputation but where this can be compensated for by better visible job performance in terms of processing times. I then derive predictions and test these against the results of my experiment.

## **5.1 Do Faster Processing Times Decrease Bribe Payments?**

The results in Section 4 are not consistent with theories of a causal relationship between faster average processing times and lower bribe payments. While the scorecards did reduce processing times, it did not reduce bribes, as shown in Tables I.2 and I.5, respectively. This is true even for the offices that were under-performing at baseline and improved their processing time the most, as shown in Tables I.6 and I.7.

One potential reason for the lack of effect from the scorecards on bribe payments could be that the information about the improvement in processing times had not yet disseminated among applicants. There are two reasons why this is not plausible. First, the scorecards did decrease expected processing times. Column (1) of Appendix Table I.A.11 shows the effect of the scorecards on the expected total processing time at the time of the first survey interview. The scorecards reduce expected processing times by 9%, similar in magnitude to the effect on actual processing times and statistically significant. Second, to further rule out that the lack of information about the improved processing times limits the effect of the scorecards on bribes, I use the information treatment that was designed to inform applicants about improvements in the processing times, as described in Section 2.7. Column (2) of Appendix Table I.A.11 shows that the point estimate for the effect of the information intervention on expected processing times is a reduction of 4% but that the estimate is not statistically significant. Taken together, these results suggest that applicants are aware of the current processing times in their sub-district land office and that providing them with more information does not substantially change their expectations. Appendix Section C.4 shows that the information treatment did not affect bribes, neither by itself nor in combination with the scorecards.

## 5.2 Implications for Other Theories

Given the positive effect of the scorecards on bribes, is it possible that faster average processing times lead to higher bribes? Tables I.6 and I.7 show that for offices under-performing at baseline scorecards improved processing times the most but did not change bribe payments. This is inconsistent with any model where average processing times has a causal effect on bribe payments. Furthermore, for the offices over-performing at baseline, the scorecards increased bribe payments without changing the processing times, which is inconsistent with models of a causal effect of bribes on processing times.

The increase in bribes among the offices that were over-performing at baseline is also inconsistent with models where it is an applicants' outside option or ability to pay that determine the bribe levels (Svensson, 2003; Niehaus and Sukhtankar, 2013b). If bribe levels change as a result of a positive scorecard sent to the government official responsible for the service for which the bribe is paid, without any observable change in service quality. The bribe level cannot be fully determined by the applicants' outside option or ability to pay. This result is most likely dependent on the structure of the interaction in which the bribe is paid. In this context, the land office is the only institution that can make the required land record change and there are no close substitutes to this service. Therefore, the bribe level is expected to be determined mainly by other factors. If there had been competition for applicants between land offices, or a close alternative to a land record change, it is plausible that these outside options (or "exit" options) would have been more important in determining the bribe level (Svensson, 2003).

## 6 Model of Bureaucrat Behavior

In this section, I will provide an overview of the model I propose to explain the results of the experiment. Appendix Section A.1 provides a formal presentation of the model.

## 6.1 Model Set-up

In the model, bureaucrats get utility from a reputational concerns term which is a function of visible job performance in terms of delays and bribe money.<sup>28</sup> Bureaucrats get disutility from effort, but effort is needed to avoid delays, which decrease the reputational concerns term. Reputational concerns has decreasing marginal utility, and so does bribe money, while effort has increasing marginal disutility. Bribes and delays both reduce the reputational concerns term since if a bureaucrat consistently asks for high bribes and do not process applications on time, a negative reputation about the bureaucrat is built and becomes visible to others. The visibility of delays increases with the scorecards, making delays more important for the bureaucrats' reputational concerns.

Bureaucrats differ only in the extent to which they care about their reputational concerns. This could be because of differences in discounting future career prospects, differences in the valuation of social status from holding a high-level civil service position, or differences in intrinsic motivation to "do a good job." What is important about these differences for the model is that they create the difference between over- and under-performing bureaucrats.<sup>29</sup> This assumption is also consistent with the observation that over-performing bureaucrats collect less bribes than under-performing bureaucrats in the control group. This would not be the case if ability is what made over-performing bureaucrats better than under-performing bureaucrats.<sup>30</sup>

In the model, the applicants simply pay the bribe amount that the bureaucrats are demanding. While this is clearly an abstraction from reality, Figure I.1 shows that the average stated value of a record of rights for applicants is substantially higher than the values of bribes paid. Even the largest estimate for the average bribe, the estimate of a typical payment, is just 0.1% of the average estimated value of the record of rights. This difference between the applicant valuation and the amount paid suggests that the applicants' willingness to pay for the service is not an important

---

<sup>28</sup>The reputational concerns term represents reasons for why the bureaucrats cares about what others, especially their supervisors, think of them. This could be for material reasons, such as career progression, social reasons, such as maintaining a good social standing with others in the bureaucracy, or psychological reasons such as the negative feelings of pride (or shame) stemming from knowing that someone else knows about one's good (or bad) performance. The term also encapsulates psychological reasons that are internal to the bureaucrats, such as the negative feelings stemming from failing to perform one's duty or breaking an internalized social norm of performing at least as well as one's peers. I.e., the effect of the scorecards on the reputational concerns term captures both possible mechanisms described in Section 4.2.

<sup>29</sup>Ashraf et al. (2020) show that differences in the motivations of public servants is important for public service delivery.

<sup>30</sup>Table I.7 shows that in the control group, offices under-performing at baseline also extract substantially larger bribes. This would not be the relationship if the differences in processing times were driven by a bureaucrat characteristic uncorrelated with bribe payments, such as ability. However, this relationship should be interpreted as an association, as the causal effect of the bureaucrat type on bribes is not identified by the experiment.

determinant of the bribe value. Instead, what determines the amount of bribes that the bureaucrats extract in the model is the trade-off between bribe money and reputational concerns.

The model also abstracts away from bribes that increase the speed of processing and the value that fast processing-times have to the applicants. While this assumption is a simplification, Figure I.1 shows that the average value an applicant put, even on the fastest reasonable processing time, is just 33% of the average estimated value of a typical bribe payment. This suggests that most of the bribes are not paid for increasing the speed of processing.

Finally, the model assumes that bureaucrats cannot buy reputation using money. This abstracts away from situations where bureaucrats use bribe money to pay supervisors for promotions, but the results would be the same if bureaucrats would pay for the position as ACL in the first place but that they then cannot bribe their way to future career advancement or high social standing in the bureaucracy.<sup>31</sup>

## 6.2 Model Predictions

The theoretical model has two main testable predictions. In what follows, I describe these predictions, the intuition behind them, and how I test them empirically. Appendix Section A.1 provides the formal model as well as the derivations and formal statements of the predictions.

### 6.2.1 Effects of Scorecard on Delays

The first set of predictions relates to the effect of scorecards on delays. The scorecards have two different effects on delays, a *substitution effect* and an *income effect*. These effects are analogous to the substitution and income effects from a wage increase in a standard labor supply model. The substitution effect leads to a decrease in delays for all bureaucrats. This is because the scorecards increase the importance of delays for bureaucrats' reputational concerns. Therefore, the marginal effect on utility from decreasing delays increases and bureaucrats provide more effort to avoid delays.

The income effect from the scorecards on delays is positive for over-performing bureaucrats and negative for under-performing bureaucrats. For over-performing bureaucrats, the scorecards

---

<sup>31</sup>Weaver (2020) analyses the effects of such bribes in the allocation of job applicants to positions in public service delivery.

increase their reputation by making the already positive performance more visible. Since reputation has decreasing marginal utility, this decreases the marginal utility effect from changes to their reputation and reduces the optimal amount of effort they provide to avoid delays. Therefore, the model does not have a prediction for the effect of the scorecards on delays among over-performing bureaucrats, the direction of the effect depend on if the substitution or income effects is stronger. For under-performing bureaucrats, the scorecards make the negative performance more visible and make their reputation worse. This increases the marginal utility from reputation and hence increases the optimal amount of effort that bureaucrats provide to avoid delays. Hence, for these bureaucrats the substitution effect and income effect are in the same direction and the model predicts that the scorecards will reduce delays among under-performing bureaucrats.

Prediction 1:      Scorecards improve processing times for bureaucrats under-performing at baseline

Inconclusive:      Ambiguous direction of the effect on processing times for bureaucrats over-performing at baseline

These predictions are tested directly in Table I.6, described in Section 4.5. Consistent with Prediction 1, the scorecards improve processing times for offices under-performing at baseline leading to fewer delays and shorter average processing times. The effect for over-performing offices is substantially smaller than the effect for under-performers and the difference in the two effects on the ICW index is marginally statistically significant.

### **6.2.2 Effects of Scorecard on Bribes**

The second prediction relates to the effect of scorecards on bribe amounts extracted from applicants and the levels of these amounts in the control group. In the model, the bureaucrats could increase bribes by simply asking applicants for more money to approve their applications. What constraints bureaucrats from extracting more bribes is their reputational concerns. Therefore, the marginal utility from reputation is an important determining factor for bribe payments. When the scorecards improve over-performing bureaucrats' reputation, the marginal negative effect bribes have on utility through the reputational concerns channel declines. This leads to an increase in bribes taken

by over-performing bureaucrats when they receive the scorecards.<sup>32</sup>

For under-performing bureaucrats, the decrease in the reputational concerns term leads to an increase in marginal disutility from bribes coming through the reputation channel. This could lead to a decrease in bribes, but since effort increases in response to the scorecards, the overall effect on reputation could be positive or negative. Since the effect on bribes for this group is ambiguous, the model does not have a prediction for the effect of the scorecards on bribes for under-performing bureaucrats.

Prediction 2:      Scorecards increase bribes for bureaucrats over-performing at baseline

Inconclusive:      Ambiguous direction for the effect on bureaucrats under-performing at baseline

These predictions are tested directly in Table I.7, described in Section 4.5.2. The scorecards increase bribes paid in offices over-performing at baseline. The effect of the scorecards on bribes is close to zero for office under-performing at baseline.

### **6.3 Potential General Equilibrium Effects within the Civil Service**

In the predictions described above, I do not allow for the scorecards to change the benchmark performance that bureaucrats are compared against. In the context of the experiment, this does not qualitatively alter the predictions since half of the bureaucrats creating the benchmark do not receive the scorecards. However, if the scorecards were to be scaled-up to all bureaucrats, there would be a larger effect on the benchmark performance. This would shift the whole distribution of performance percentiles down and thereby have an income effect on all bureaucrats. The prediction from the model is that this income effect would induce more effort and smaller bribe payments than the partial experimental roll-out of the scorecards.

---

<sup>32</sup>This effect is dependent on that the income effect on delays is not so strong that it dominates the substitution effect and mutes any positive effect on the reputational concerns coming from the increased visibility of the already good performance. We see that this is not the case in Table I.6, where the overall performance of bureaucrats over-performing at baseline is marginally positive, suggesting that the substitution effect marginally dominates the income effect and hence the scorecards lead to an increase in the reputation of over-performing bureaucrats by increasing the visibility of their positive performance.

## **6.4 Alternative Explanations for the Effects of Scorecards**

### **6.4.1 Increased Marginal Costs of Bureaucrats' Time or Increased Willingness to Pay Among Applicants**

One potential explanation for the scorecards causing higher bribe payments, is that scorecards increase the marginal value of the bureaucrats' time. If bribe payments are made so that the bureaucrats spend more time on an application, the marginal value of the bureaucrats' time could be an important determinant of the bribe amount. If the scorecards increase the overall amount of time that bureaucrats are working, it is also likely that the marginal value of their time increased. Hence, it is possible for scorecards to have increased bribes through this mechanism. Another alternative explanation is that faster processing times lead applicants to be willing to pay more to get their land record change.

However, both of these explanations are inconsistent with the result that in the offices where the changes in processing times were the largest, bribe payments did not change. Instead, it is in the offices where changes in processing times are small that bribe payments increase. If it had been an increase in the willingness to pay by applicants or an increase in bureaucrats' effort that lead to the increase in bribes, the increase would take place in the offices that were under-performing at baseline, because these are the offices where the scorecards improve processing times. Therefore, it is unlikely that either of these mechanisms is a substantial reason for the increase in bribe payments.

### **6.4.2 Transfers of Over-performing Bureaucrats**

An alternative explanation, that is consistent with the heterogeneity in the effects on delays and bribes, is that over-performing bureaucrats get transferred due to receiving positive scorecards and that they are replaced by average performing bureaucrats. If the average performing bureaucrats both have slower processing times and collect more bribes, we expect that bribe payments would increase in offices over-performing at baseline. Processing times may not change as the incentive effects of the scorecards may cancel out the effect of high quality bureaucrats being replaced by lower quality bureaucrats.

However, this explanation is refuted by the data on bureaucrat transfers. Appendix Table I.A.12 shows that the scorecards did not affect bureaucrats' transfers. Column (1) shows the effect on

the monthly probability of being transferred, Column (2) shows the heterogeneity in the effect by offices over-performing and under-performing at baseline. Columns (3) and (4) show the overall and heterogeneous effects on the duration of the posting for the first bureaucrat after the start of the experiment, including postings that started before the experiment. Columns (5) and (6) show the overall and heterogeneous effects on not having any ACL assigned to the office. All of the effects are close to zero and not statistically significant.

#### **6.4.3 Over-performing Bureaucrats using Scorecards in Negotiations over Bribes with Applicants**

Another alternative explanation is that positive scorecards help bureaucrats prove to applicants that they have the ability to process applications quickly. This could then allow the bureaucrats receiving positive scorecards to charge higher bribes while it would not affect the bribes in offices receiving negative scorecards since these would not be shown to applicants.

There are three reasons why this explanation is implausible. First, the coefficients on "Overperform baseline" in Columns (4) and (5) of Appendix Table I.A.11, show that the expected processing times are 14% lower in the over-performing offices not receiving the scorecards, suggesting that the applicants are already aware of the faster processing times in these offices. Furthermore, in Column (4) the point estimate for how the scorecards effect on applicants' expectations in over-performing offices is a 6% decrease, similar to the point estimate of a 3% decrease for the actual improvement of the processing times in these offices, as shown in Column (2) of Table I.4. If the scorecards helped bureaucrats change applicants' expectations, the effect on the expectations should be larger than the effect on the actual processing times. Second, although I cannot rigorously rule out that no one in the land offices showed the scorecards to applicants, in none of the qualitative interviews done with ALCs and applicants was it even mentioned that the scorecards were shown to applicants and when directly asked, the applicants said they were not aware of the performance scorecards. Third, the information intervention tried to accomplish the effect that a bureaucrat could achieve by showing the scorecard to an applicant. Column (2) in Appendix Table I.A.11 shows that the point estimate of the effect of the information intervention is just a 4% improvement in the expected processing time, suggesting that it is difficult to move applicants priors through simple information



interventions. Furthermore, Appendix Table I.A.4 shows that the information intervention did not increase bribes.

## 7 Conclusion

I have shown that information flows about individual government bureaucrats performance within a bureaucracy can improve the performance of these bureaucrats, even in the absence of explicit performance incentives. The results from the experiment show that these effects can happen rapidly and persist over at least 16 months. One plausible mechanism for this effect is that the bureaucrats care about the reputation they have among their supervisors. A second potential mechanism is that being measured and compared to your peers increases the salience of the performance to the bureaucrats themselves and generates a sense of shame or pride that create an additional motivation to perform well.

One way to assess the value of the improved processing times is to multiply the applicants' average stated valuation of having their application processed one day faster with the reduction in the total number of processing days due to the scorecards.<sup>33</sup> For the 155 offices receiving the treatment, this gives a value of approximately USD 9.7 million per year.<sup>34</sup> This value should be interpreted carefully since it relies heavily on the stated value of faster processing to the applicants. However, the number is more than two orders of magnitude larger than the implementation costs of the scorecards, which were approximately USD 20,000 per year, even when including the author's time and set-up costs. However, the value of the experimental intervention becomes less clear when taking into account the effect on bribes. Multiplying the effect of the scorecards on reported payments with the number of applications in the treatment area results in an estimate of the effect on total bribes paid of 1.9 million per year. If the effect on the estimated typical payment is used instead, the total increase is USD 7.6 million per year.

Except for the increase in bribe payments, I do not find any evidence for unintended consequences or gaming of the scorecard's quantitative performance indicators. It is possible that

---

<sup>33</sup>I calculate the value of having the application being processed one day faster using the following formula:  $\frac{\text{Value of processing in 7 days}}{\text{Expected processing time from survey date} - 7}$ . All the information comes from the in-person survey made before the application was actually processed.

<sup>34</sup>The number of applications per year is estimated by taking the number of applications in the last six months of 2019 when all offices had the e-governance system installed and multiplying by 2.

monitoring or information flows that are not directly tied to explicit incentives are less likely to have the unintended consequences that are common for explicit incentive structures. One reason for this is that the receivers of the information can interpret the information flexibly. If bureaucrats engaged in observable behavior leading to unintended consequences, it would be possible for the supervisors to take this into account when interpreting the information on the scorecards. Furthermore, the scorecards were well received by most supervisors and there was no substantial backlash among bureaucrats. This points to another difference from explicit performance incentives. Since improved information flows do not reduce the discretionary power of supervisors, it is possible that they are less likely to be opposed by important actors in the organization and, therefore, poorly implemented (Banerjee et al., 2020).

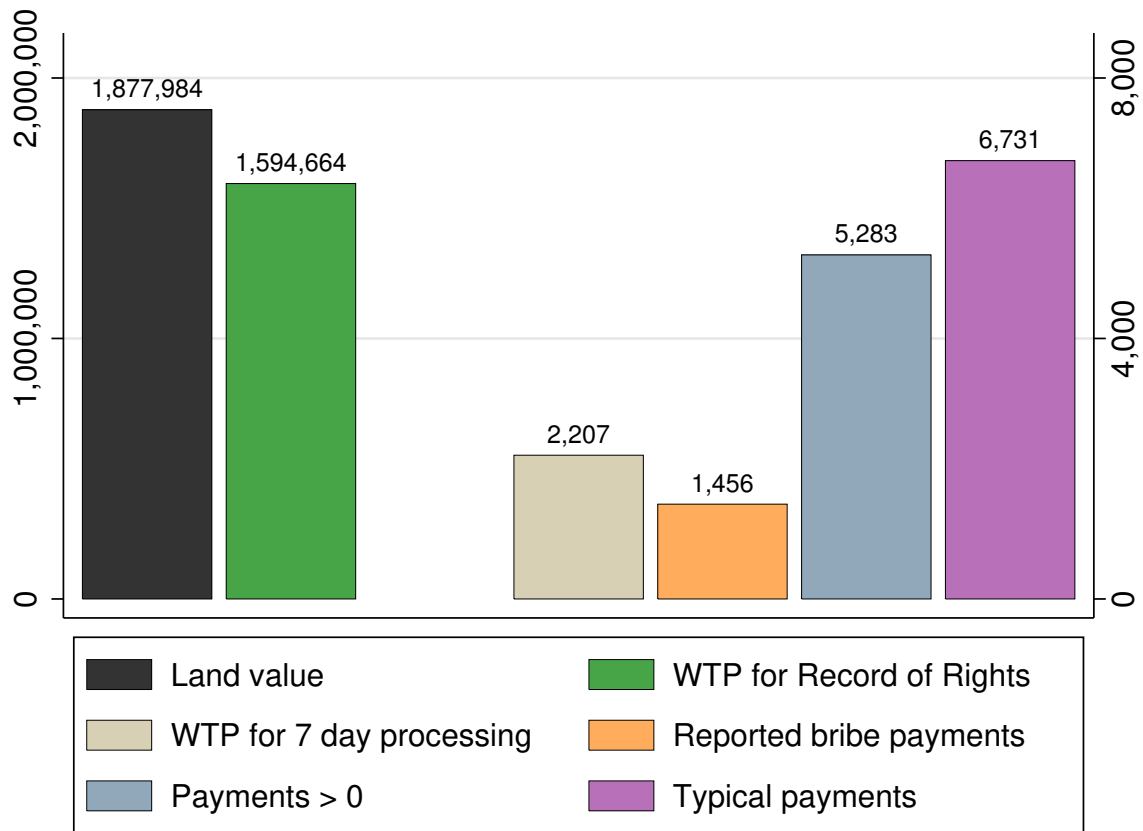
The results have several policy implications. First, the result highlights that there exists untapped potential in data generated by e-governance systems. As more and more public services are delivered using e-governance systems, the cost of monitoring and evaluating civil servants' performance has drastically decreased. The results of the experiment show that using the data generated by e-governance systems for monitoring and evaluation has significant potential to improve bureaucratic efficiency.

Second, the differential effects of the scorecards on under-performing and over-performing offices suggest that it is especially important to improve information flows for under-performing bureaucrats. Regardless of the mechanism for this result, it implies that the type of recognition systems that are common for bureaucrats in low- and middle-income countries, where outstanding performances are recognized without addressing inadequate performances, are ineffective. This is because providing positive feedback has a negative effect stemming from the improved reputation that the positive performance information generates. Instead, it is more important to make sure negative feedback is provided to under-performing civil servants. Positive feedback might still have an overall positive effect since it may motivate under-performing bureaucrats who want to receive better feedback, but the positive feedback is likely less effective than the negative feedback and can, in some cases, even be counter-productive.

Finally, the model points out a more general problem when using reputational concerns to incentivize a socially desirable behavior by an agent. Any reform or intervention that increases the reputation of some agents may also have a negative spill-over on other behaviors where reputation

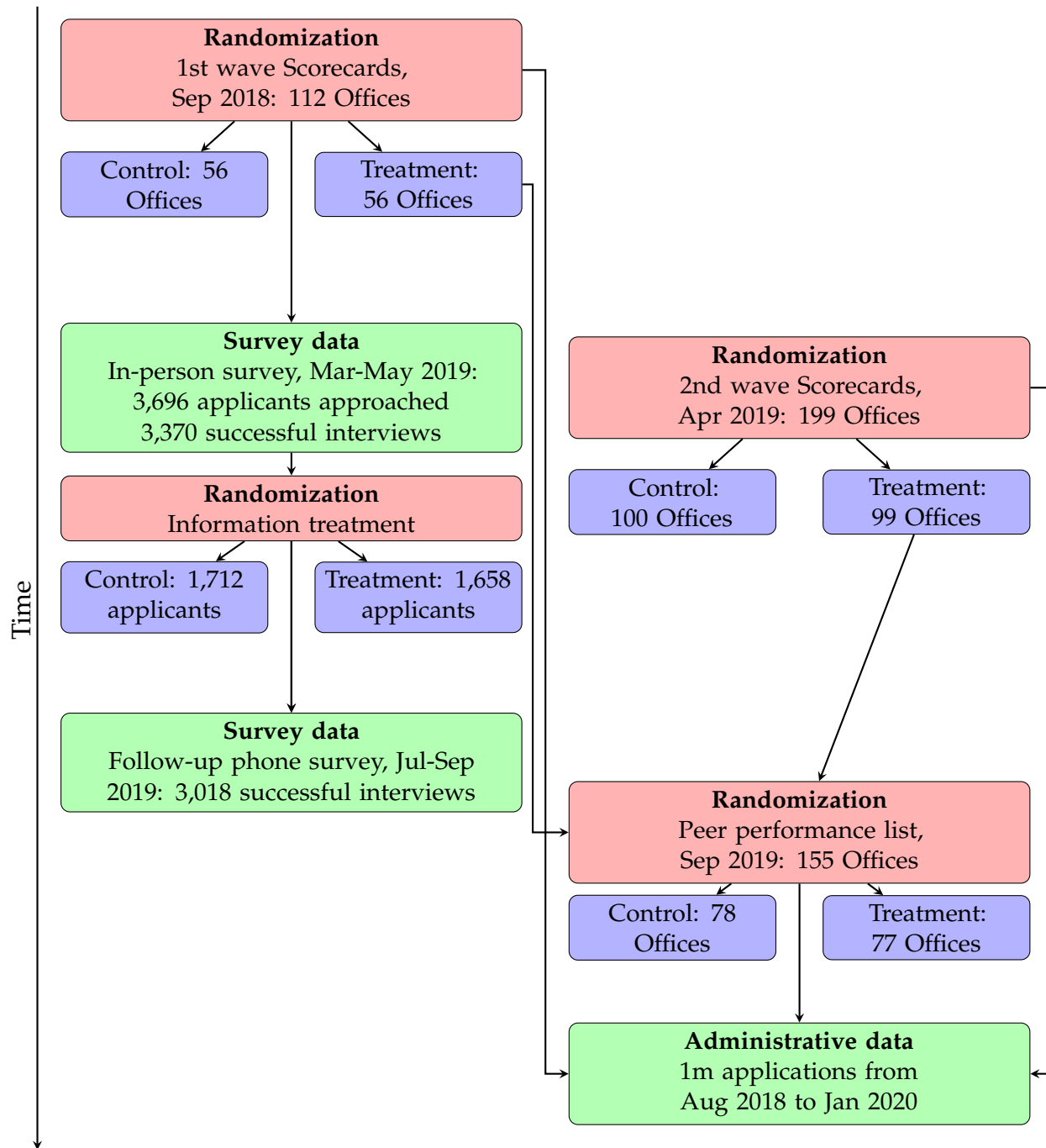
is a motivating factor. This is an especially important insight for government bureaucracies, where compressed wage structures and secure employment of civil servants often make reputational concerns more important motivators than in other organizations.

Figure I.1: Value of Land, Record of Rights, Faster Processing, and Bribe Payments



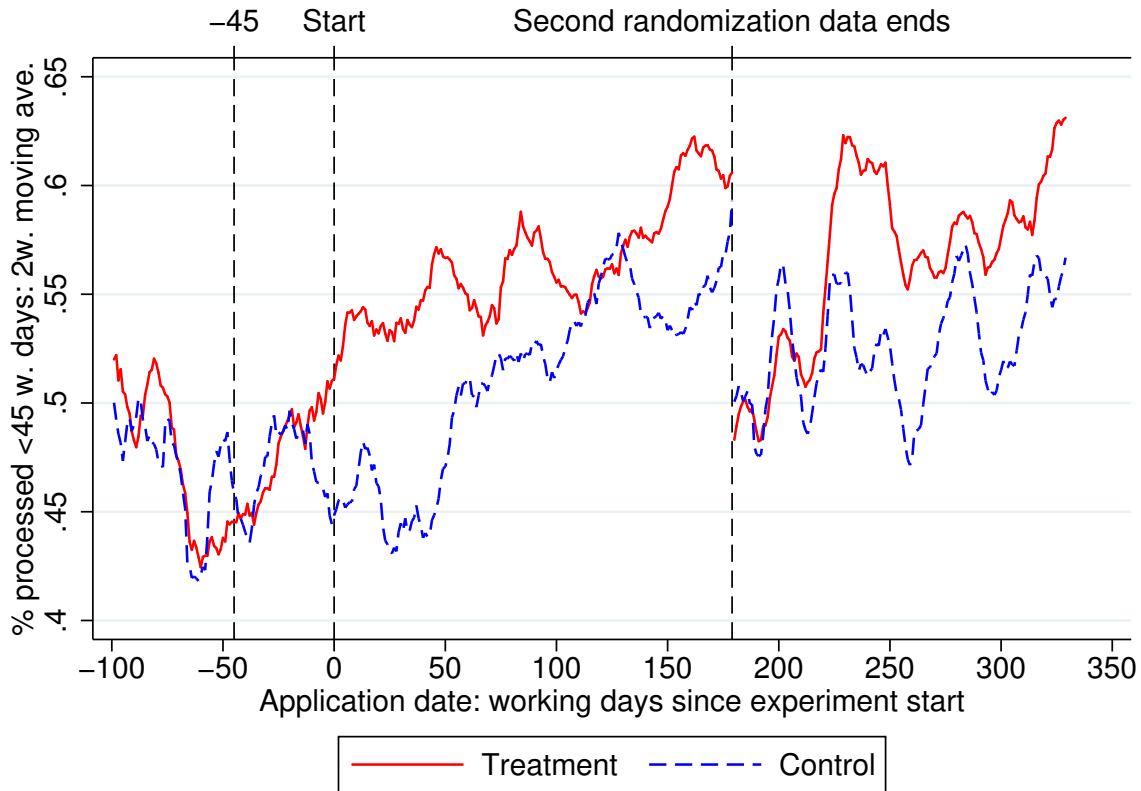
The first bar of this figure shows the average of applicants' estimates of the value of the land for which the land record change is being applied for. The second bar shows the average of applicants' stated value of getting the land record change approved and receiving a record of rights. The third bar shows the average stated value of getting the application processed within seven days from the time of the first survey. The fourth bar shows the average value of bribe payments reported by the applicant, 73% of the applicants reported having paid no bribes. The fifth bar shows the average value of bribe payments reported by applicants reporting having paid some bribe. The sixth bar shows the average value of an estimated "typical bribe payment by a person like yourself" reported by the applicant, 27% of the applicants responding to this question reported that a typical applicant paid no bribes. The first two bars are measured on the axis on the left, the next four bars are measured on the axis to the right. All variables are winsorized at the 99th percentile. Observations are weighted by the inverse of the number of observations in their land office. Discussed in Sections 2.1 and 6.

Figure I.2: Overview of randomization and data collection



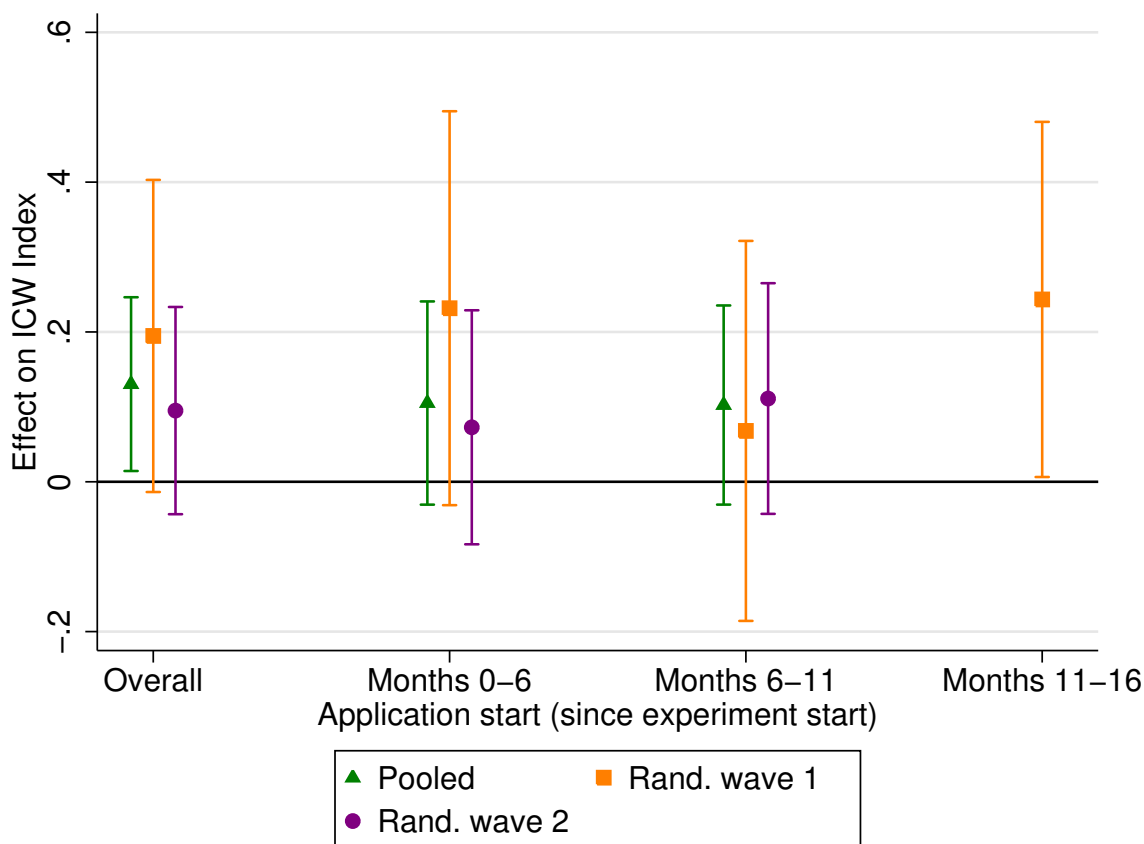
This figure provides a visual overview of the experiment design and data collection. Boxes placed further down in the figure represent things that happened later with the exception of the administrative data collection, which happened throughout the project. Red boxes represent randomizations into treatment or control. Blue boxes represent the treatment and control groups. Green boxes represent data collection. Discussed in Section 2.4.

Figure I.3: Fraction of Applications Processed within 45 Working Days



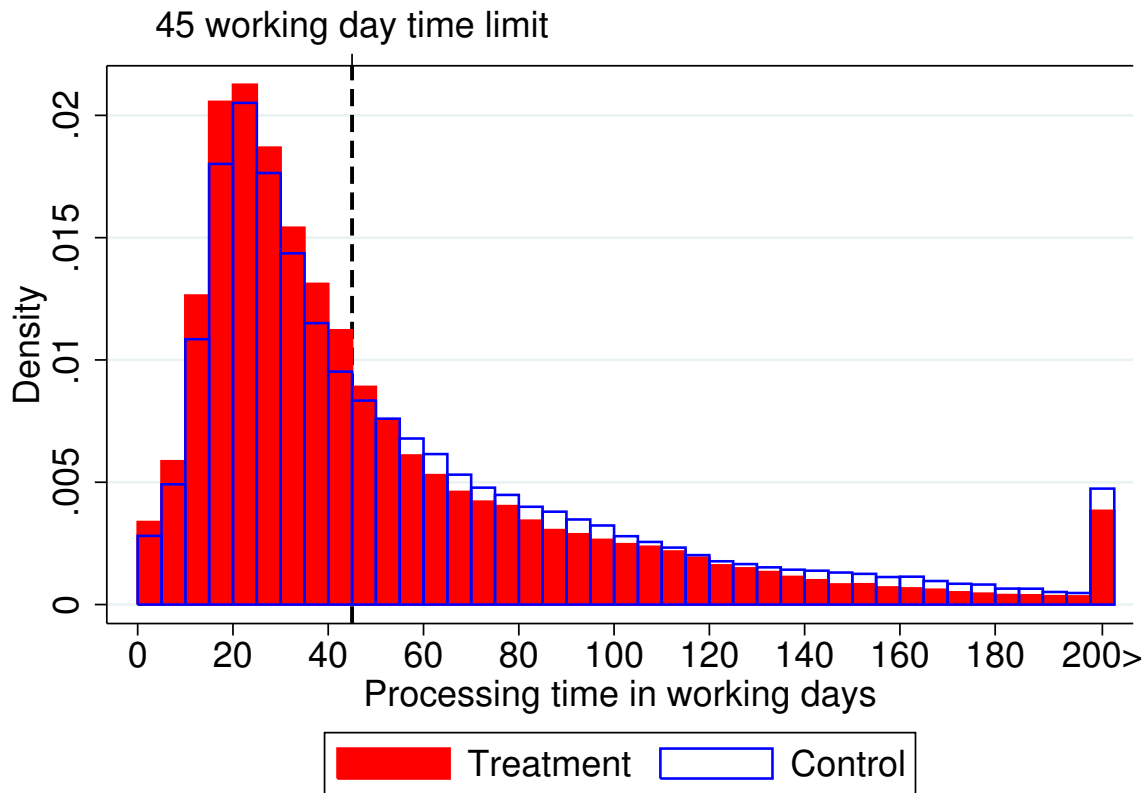
This figure shows the two week moving average of the daily fraction of applications processed within the 45 working day limit in the treatment and control groups. Data contains all applications made from 100 working days before the start of the experiment until 45 working days before the experiment ended (25 Apr 2018 - 20 Jan 2020). The first vertical line represents the date 45 working days before the first scorecard was sent out. In principle, the effect can have started for any application made after this date. The second vertical line represents the date of the first scorecard, applications made after this date were fully treated. The third vertical line represents the end of the data from the second randomization wave. To the right of this line the figure is based on the 112 offices in the first randomization wave. Discussed in Section 4.1.1.

Figure I.4: Effect of Scorecards by Time Since the Start of Experiment



This figure shows the regression coefficients and confidence intervals for regressions using applications started during different time periods after the experiment started. The outcome variable is the ICW Index from Column (3) of Table I.2. The overall estimate uses the same specification as in Table I.2. The estimates for the three time periods are estimated by interacting a dummy variable for if the application was made in that time period with the treatment variable. The figure shows results from regressions using data from all offices (triangle), offices in the first randomization wave (squares), and offices in the second randomization wave (circles). The months are numbered relative to when scorecards were sent out for that office's randomization wave. Month 0 is the month before the start of the experiment. Confidence intervals are constructed using standard errors clustered at the office level. Discussed in Section 4.1.1.

Figure I.5: Histogram of Processing Times by Treatment



This figure shows histograms of processing times for the treatment and control groups separately. Processing times are top coded at 200 working days. Data contains all applications made between one month before the start of the experiment and 45 working days before the experiment ended (13 Aug 2018 - 20 Jan 2020). Applications not processed as of October 2020 are excluded. Discussed in Section 4.1.2.



Table I.1: Summary Statistics

	(1) Mean	(2) Median	(3) St. Dev.	(4) Observations
<b>Panel A: Application level administrative data</b>				
Process time < 45 w. days	0.59	1	0.49	1,050,924
Actual process times (w. days)	50	34	45	972,589
Process time inc. imputed values (w. days)	63	36	70	1,050,924
Approval rate	0.69	1	0.46	972,582
<b>Panel B: Monthly office level administrative data</b>				
Total applications	287	213	272	4,516
Applications processed	240	136	331	4,516
Apps. disposed within 45 w. days	150	79	195	4,516
Apps. pending beyond 45 w. days	382	97	732	4,516
<b>Panel C: Applicant survey data</b>				
Applicant age	47	47	14	2,903
Female	0.06	0	0.24	3,018
Applicant monthly income (BDT)	23,902	20,000	21,093	2,791
Applicant HH per capita expenditure (BDT)	4,400	3,462	3,537	3,018
Land Value (BDT 100,000)	19	8	30	2,800
Land Size (Decimal = 1/100th Acre)	24	10	40	2,892
Any additional payment made	0.28	0	0.45	3,018
Reported payment amount (BDT)	1,456	0	3,456	3,018
Typical payment amount (BDT)	6,731	5,000	8,414	1,896

This table shows summary statistics for applications in the administrative data, offices, and applicants in the survey data. Observations in Panel A and C are inversely weighted by the number of applications in their land office. Observations in Panel B are uniformly weighted. Continuous variables in the survey data are winsorized at the 99th percentile. USD/BDT $\approx$ 84.3. Reported payment amount is any payment reported by the applicant above the official fee. Typical payment amount is the answer to the question of my much a "normal person, like yourself" typically pay to get an application processed. Discussed in Section 2.5.

Table I.2: Effect of Scorecards on Processing Times

	(1)	(2)	(3)
	<45 w. days	IHS(w. days)	ICW index
Scorecard	0.0608** (0.0275)	-0.125** (0.0593)	0.130** (0.0592)
Start month FE	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes
Observations	1,050,924	1,050,924	1,050,924
Clusters	311	311	311
Control mean	0.56	65.64	-0.00
Fraction imputed		0.06	
Fraction zero		0.003	

This table shows the effect of the scorecards on the speed of application processing. Column (1) shows the effect on the fraction of applications processed within the 45 working day time limit. Column (2) shows the effect on the IHS transformation of processing time. Applications that are not yet processed are given imputed processing times equal to the mean of processing times that are longer than the application has currently been pending. Column (3) shows the effect on an inverse covariance weighted matrix combining the outcome variables of Columns (1) and (2). Data contains all applications made between one month before the start of the experiment and 45 working days before the experiment ended (13 Aug 2018 - 20 Jan 2020). Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in their land office.

\*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section 4.1.

Table I.3: Effect of Peer Performance List

	(1)	(2)	(3)	(4)
	<45 w. days	IHS(w. days)	<45 w. days	IHS(w. days)
Peer Performance List	0.00483 (0.0452)	-0.0360 (0.0948)	0.0116 (0.0394)	-0.0404 (0.0823)
Scorecard			0.0570* (0.0293)	-0.112* (0.0651)
Start month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	286,152	286,152	1,050,924	1,050,924
Clusters	155	155	311	311
Control mean	0.67	51.31	0.59	65.64
Fraction imputed		0.06		0.06
Fraction zero		0.003		0.003

This table shows the effect of adding the list of performances to the scorecard and sharing the performance information about one office with the ACLs, UNOs, and DCs of 76 other offices. Column (1) shows the effect on the fraction of applications processed within the 45 working day time limit. Column (2) shows the effect on the IHS transformation of the processing time. Columns (3) and (4) show the effects of both the scorecard treatment without the performance list and the performance list separately. Columns (1) and (2) only use data from offices in the performance list experiment and data from applications made one month before the first performance list until 20 January 2020. Columns (3) and (4) use data on all applications made between one month before the start of the experiment started and 45 working days before the experiment ended (13 Aug 2018 - 20 Jan 2020). The dummy variable "Peer performance list" takes the value of one for applications made in offices receiving the peer performance lists, made later than one calendar month before the first performance list was sent out. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section 4.2.

Table I.4: Effect on Visits and Time Spent by Applicants

	(1)	(2)	(3)
	Visits	Hours spent	ICW index
Scorecard	-1.034** (0.497)	-1.586 (1.855)	0.0851 (0.0552)
Start month FE	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes
Observations	3,018	3,018	3,018
Clusters	112	112	112
Control mean	8.99	23.66	

This table shows the effect of the scorecards on visits to land offices and the number of hours spent on these visits. Standard errors are clustered at the land office level. Observations inversely weighted by the number of applications in the land office. \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ . Discussed in Section 4.3.

Table I.5: Effect on Bribe Payments for Application Processing

	Amount		Any bribe		Amount if > 0	
	(1)	(2)	(3)	(4)	(5)	(6)
Scorecard	1,046*	265	-0.014	-0.003	1,573**	1,069**
	(615)	(181)	(0.022)	(0.022)	(768)	(457)
Start month FE	Yes	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,896	3,018	1,896	3,018	1,392	807
Clusters	112	112	112	112	112	111
Control mean	6,083	1,278	0.75	0.27	8,083	4,700
Bribe measure	Typical	Reported	Typical	Reported	Typical	Reported

This table shows the effect of the scorecards on bribe payments made for application processing. Column (1) shows the effect on the estimate for how much a "normal person, like yourself" pays in bribes to process an application. Column (2) shows the effect on reported payments to government officials or agents beyond the official fee. Columns (3) and (4) show the effect on the fraction of non-zero answers for the two questions. Columns (5) and (6) show the effect among applicants who reported a non-zero bribe. All monetary amounts are in BDT. USD/BDT $\approx$ 84.3. All continuous variables are winsorized at the 99th percentile. Standard errors are clustered at the office level. Observations inversely weighted by the number of applications in the land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section 4.4.

Table I.6: Effects on Processing Times, Visits, and Time Spent by Office  
Baseline Performance

	(1)	(2)	(3)	(4)
	<45 w. days	IHS(w. days)	Office visits	Hours spent
Scorecard x Overperform baseline	0.00823 (0.0372)	-0.0345 (0.0804)	-0.678 (0.729)	0.358 (2.626)
Scorecard x Underperform baseline	0.124*** (0.0402)	-0.234*** (0.0876)	-1.230* (0.731)	-2.973 (2.843)
Overperform baseline	0.193*** (0.0502)	-0.315*** (0.108)	-1.854* (0.949)	-7.095** (3.318)
P-value sub-group diff.	0.04	0.10	0.10	0.30
Start month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	1,050,924	1,050,924	3,018	3,018
Clusters	311	311	112	112
Overperformers: q-value	1.00	1.00	0.55	1.00
Underperformers: q-value	0.02	0.03	0.24	0.55
Overperformers: control mean	0.68	51.53	8.03	20.37
Underperformers: control mean	0.41	82.54	9.88	26.72

This table shows the effect of the scorecards separately for offices with above- and below-median performance at baseline. In Columns (1) and (2), results are based on administrative data. In Columns (3) and (4), results are based on survey data. Column (1) shows the effects on the fraction of applications processed within the 45 working day limit. Column (2) shows the effects on the IHS transformation of the number of working days it took to process the application. Column (3) shows the effect on the number of visits to land offices needed for the processing of the application. Column (4) shows the effect on the number of hours spent by the applicant for the processing of the application. Standard errors are clustered at the land office level. Q-values are sharpened false discovery rate q-values for the eight hypotheses that the effect of the scorecards is zero for all outcome variables and for both over-performers and under-performers (Benjamini et al., 2006; Anderson, 2008). Observations inversely weighted by the number of applications in the land office.

\*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section 4.5.1.

Table I.7: Effect on Bribes by Office Baseline Performance

	(1) Typical payment	(2) Reported payment
Scorecard x Overperform	2279.7*** (772.9)	638.4*** (229.1)
Scorecard x Underperform	-162.4 (954.6)	-59.38 (257.8)
Overperform baseline	-1811.8* (928.4)	-819.0*** (287.8)
P-value sub-group diff.	0.06	0.05
Start month FE	Yes	Yes
Stratum FE	Yes	Yes
Weighted by office	Yes	Yes
Observations	1,896	3,018
Clusters	112	112
Overperformers: q-value	0.01	0.01
Underperformers: q-value	0.76	0.76
Overperformers: control mean	5,313	916
Underperformers: control mean	6,817	1,616

This table shows the effect of the scorecards on offices with above- or below-median performance at baseline. Column (1) shows the effects on what the applicant reports to be a typical payment for a land record change for a person like themselves. Column (2) shows the effects on the payments reported by the applicant. The outcome variables are in BDT. USD/BDT $\approx$ 84.3. Standard errors are clustered at the land office level. Q-values are sharpened false discovery rate q-values for the four hypotheses that the effect of the scorecards is zero on both outcome variables and for both over-performers and under-performers (Benjamini et al., 2006; Anderson, 2008). Observations are inversely weighted by the number of applications in that land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section 4.5.2.

# Appendices

## A Theory

### A.1 Model of Reputation and Bureaucrat Behavior

#### A.1.1 Model Set-up

Government bureaucrats utility function:

$$U \left( E_i, B_i, t_i R(B_i, v^{T_i} P(E_i)) \right) = D(E_i) + M(B_i) + t_i R \left( B_i, v^{T_i} P(E_i) \right) \quad (3)$$

- Subscript  $i$  represent individual bureaucrats
- $D(E)$  is disutility of effort  $E$ , which has negative first and second derivatives,  $D'(E) < 0$  and  $D''(E) \leq 0$
- $M(B)$  is the utility of bribe money  $B$ , which has a positive first derivative and a negative second derivative,  $M'(B) > 0$  and  $M''(B) \leq 0$
- $tR(B, vP(E))$  is the utility from reputational concerns which is determined by  $B$  and visible performance  $vP(E)$ 
  - $R(\cdot)$  has a negative derivative with respect to bribes  $R^1(B, vP(E)) < 0$  and a positive derivative with respect to visible performance  $R^2(B, vP(E)) > 0$ . Both second derivatives are negative,  $R^{11}(B, vP(E)) < 0$  and  $R^{22}(B, vP(E)) < 0$
  - The cross derivative is positive,  $R^{21}(B, vP(E)) > 0$ , i.e., bribes and performance are complements, or equivalently honesty (the lack of bribes) and performance are substitutes
  - A technical assumption used to ensure the existence of derivatives and avoid corner solutions is  $R^{12}(B, vP(E)) \leq (R^{11}(B, vP(E)) R^{22}(B, vP(E)))^{\frac{1}{2}}$
- $v^{T_i}$  is the visibility of performance  $P(E_i)$  and depend on bureaucrat  $i$ 's treatment  $T_i \in \{scorecard, control\}$  such that  $v^{scorecard} > v^{control}$



- Connecting the performance term directly to the scorecards, I assume that performance is the average ranking of a bureaucrat in terms of applications processed on time and applications pending longer than the time limit
- $P(E)$  is increasing in  $E$ , positive when  $E$  is above median effort and negative when  $E$  is below median effort
  - \* The second derivative is zero or negative,  $P''(E) \leq 0$
- $t$  is the type of bureaucrat and reflects the degree to which the bureaucrat values reputation
  - Bureaucrats only differ in their valuation of reputation  $t$  and their treatment status  $v$

All of the assumptions, including the technical assumptions mentioned below, are fulfilled by a simple Cobb-Douglas utility function of the form:

$$U = \alpha \ln(1 - E) + \beta \ln(B) + t \ln(c + v(E - \bar{E}) - B)$$

Where  $c$  is a constant sufficiently large so that  $c + v(E - \bar{E}) - B > 0$  and  $\bar{E}$  is the effort of the median bureaucrat when  $v^T = v^{control}$ .

For simplicity, I do not formally model applicants' behavior but assume that they have no choice but to accept the bureaucrats' bribe request.

### A.1.2 Solution to Bureaucrats' Problem

Bureaucrats choose  $E$  and  $B$  to maximize  $U(E, B, tC(B, vP(E)))$ . The first order conditions to the bureaucrats maximization problem are:

$$D'(E_i^*) + t_i R^2(B_i^*, v^T P(E_i^*)) v^T P'(E_i^*) = 0 \quad (4)$$

Where  $E_i^*$  and  $B_i^*$  represent the choices of  $E$  and  $B$  that maximize utility for bureaucrat  $i$ . At the optimum, the marginal disutility of effort equals the marginal utility of effort's effect on reputational concerns.

$$M'(B_i^*) + t_i R^1(B_i^*, v^T P(E_i^*)) = 0 \quad (5)$$

At the optimum, the marginal utility of bribe money equals the marginal disutility from a decrease in reputation due to bribes.

### A.1.3 Effect of Scorecards on Effort

Henceforth, I will drop the star superscript (\*) on  $E$  and  $B$  since all mentions will refer to the values of  $E$  and  $B$  at the optimum. Taking the total derivatives of the first order conditions with respect to  $v$  gives us the following expression for the derivative of  $E$  with respect to  $v$ :

$$\frac{dE_i}{dv} = P'(E_i) \frac{-\left(\frac{M''(B_i)}{t_i} + R^{11}(\cdot)\right) R^2(\cdot) + \left((R^{12}(\cdot))^2 - \left(\frac{M''(B_i)}{t_i} + R^{11}(\cdot)\right) R^{22}(\cdot)\right) v^T P(E_i)}{\left(\frac{D''(E_i)}{t_i} + R^{22}(\cdot) (v^T P'(E_i))^2 + R^2(\cdot) v^T P''(E_i)\right) \left(\frac{M''(B_i)}{t_i} + R^{11}(\cdot)\right) - (R^{12}(\cdot) v^T P'(E_i))^2} \quad (6)$$

Using the technical assumption  $R^{21}(B, vP(E)) \leq (R^{11}(B, vP(E)) R^{22}(B, vP(E)))^{\frac{1}{2}}$  the denominator is positive. The first term in the numerator reflects the direct substitution effect of the changed visibility of effort. This effect is always positive. The second term in the numerator represents the income effect, i.e., the effect on effort stemming from improved reputation because of the change in visibility of existing efforts. If  $P(E) < 0$  then this term is also positive making the whole expression positive.

Prediction 1: If  $P(E) < 0$  then  $\frac{dE}{dv} > 0$ . I.e., the scorecards improve processing times for offices under-performing at baseline

For bureaucrats with an above-median performance, when their positive performance becomes more visible, their reputation improves, and the marginal utility from exerting effort on improving reputation decreases. For bureaucrats with an above-median effort, the effect is, therefore, ambiguous.<sup>35</sup>

### A.1.4 Effect of Scorecards on Bribes

Taking the total derivatives of the first order conditions with respect to  $v$  gives us the following expression for the derivative of  $B$  with respect to  $v$ :

<sup>35</sup>The ambiguous effect is analogous to the effect of a wage increase on labor supply. The income effect and substitution effect go in different directions and depending on which dominates the overall effect may be positive or negative.

$$\frac{dB_i}{dv} = R^{12}(\cdot) \frac{v(P'(E_i))^2 R^2(\cdot) - \left( \frac{D''(E_i)}{t_i} + 2R^{22}(\cdot) (v^T P'(E_i))^2 + R^2(\cdot) vP''(E_i) \right) P(E_i)}{\left( \frac{M''(B_i)}{t_i} + R^{11}(\cdot) \right) \left( \frac{D''(E_i)}{t} + R^{22}(\cdot) (vP'(E_i))^2 + R^2(\cdot) vP''(E_i) \right) - (R^{21}(\cdot) vP'(E_i))^2} \quad (7)$$

Again using the technical assumption  $R^{21}(B, vP(E)) \leq (R^{11}(B, vP(E)) R^{22}(B, vP(E)))^{\frac{1}{2}}$  the denominator is positive. The first term in the numerator is positive and derived from the substitution effect increasing effort, which leads to an improvement in visible performance which in turn leads to an increase in bribes because of the complementarity between visible performance and bribes. The second term is derived from the changed visibility of pre-existing performance which also affects bribes due to the complementarity between visible performance and bribes. For bureaucrats with an effort above the median effort this effect is positive. Their positive performance becomes more visible so their reputation term improve and the complementarity decreases the marginal disutility, through the reputation channel, from collecting bribes. Conversely, for bureaucrats with below-median effort this effect is negative. Hence, for bureaucrats with above-median effort the scorecards leads to higher bribes ( $\frac{dB}{dv} > 0$ ) while for bureaucrats with below-median effort the effect is ambiguous.

Prediction 2: If  $P(E) > 0$  then  $\frac{dB}{dv} > 0$ . I.e., scorecards increase bribes for offices over-performing at baseline

## A.2 Monopolistic Price Discrimination Model

The experiment was designed to test a specific model of how the speed of application processing and bribes are connected. A full exposition of the complete model and its predictions is available in the pre-analysis plan. Here I outline the intuition behind the model and its prediction for the experiment. The model is based on an asymmetric information model of price discrimination under monopoly where the bureaucrat acts as monopolists selling a service. Applicants get utility from having their application processed, the faster the application is processed the more utility the processing generates. Applicants differ only in their willingness to pay for the speed of processing their applications. All applicants' utility is linear in money and none of them are liquidity constrained so their willingness to pay equals their ability to pay. The bureaucrat can ask

for different bribe payments from the applicants and can offer the service with different processing times or refuse to provide the service. Once a processing time and bribe payment is agreed upon the applicant pays the bribe and the bureaucrat must honor the agreement. The bureaucrat gets utility from receiving bribes. It can be costly for the bureaucrats to process applications faster although this is not necessary for the main conclusions of the model.

Perfect information in the context of this model means that the bureaucrat can perfectly observe the applicants' willingness to pay for having the application processed faster. Under perfect information, applicants will have their applications processed at a Pareto optimal speed where the marginal benefit of having the application processed faster is the same as the marginal cost of processing the application faster for the bureaucrat. Asymmetric information means that bureaucrats cannot observe the applicants' willingness to pay. Under asymmetric information, the bureaucrat has to offer the same menu of processing times and bribe payments to all applicants.<sup>36</sup> Under asymmetric information only the applicants with the highest willingness get their application processed at the Pareto optimal speed. All other applicants have their applications slowed down as the bureaucrats trade-off providing fast processing for applicants with lower willingness to pay with how large of a bribe they can charge from applicants with a higher willingness to pay.

An easy way to see this trade-off is in a simple example where it is costless for the bureaucrat to process the application immediately and there are two types of applicants, one with a higher willingness to pay to have the application processed quickly. Under full information, the bureaucrat can simply make a take-it-or-leave-it offer to all applicants at exactly their willingness to pay to have the application processed immediately. The applicants will pay their respective willingness to pay because they have no better outside option and the bureaucrat will process the applications immediately. Under asymmetric information, the bureaucrat cannot differentiate between the applicants ex-ante. It now becomes optimal, from the bureaucrat's perspective, to offer to process the application immediately at a higher bribe payment and slower at a lower bribe payment. The applications for those with low willingness to pay are now intentionally delayed despite that processing them immediately does not cost the bureaucrat anything.

---

<sup>36</sup>Realistically, some observable characteristics contain information about the applicants' willingness to pay. In this case, the bureaucrat has to offer the same menu to all applicants with the same observable characteristics.

### **A.2.1 Model's Predictions for the Experiment**

The scorecards encourage bureaucrats to process applications within 45 working days. Section 4 shows that the scorecards led to an increase in the applications processed within 45 working days and that the effect was driven by offices that were under-performing at the start of the experiment.

Under full information, an increase in processing speed is predicted to lead to a slight increase in bribe payments among those whose applications are processed faster. This is because for these applications the value of the processing has increased and they are now willing to pay more for it. Under asymmetric information, an increase in processing speeds is predicted to reduce the bribe payments among those with the highest willingness to pay for getting their applications processed quickly. This is because the bureaucrat has to make the menu option of having applications being processed quickly more attractive in order for these applicants to continue to pay for it now that the processing speed of the option to pay less has increased.

### **A.2.2 Testing the Model's Predictions**

The main testable prediction of this model with respect to the experiment is that when delays in processing times are reduced, bribe payments should decrease among those who have a high willingness to pay for getting their application processed quickly. This result stems from that the bureaucrat uses the difference in terms of processing times between those paying large and small bribes, to maintain a separating equilibrium and maximize the amount of bribes extracted. In particular, if long delays were reduced, but the bribes for those who paid the largest bribes and subsequently got the fastest processing time did not change, then some of these applicants would choose to pay a lower bribe and get a slower processing time. Anticipating this the bureaucrats would reduce the bribes for those with the highest willingness to pay for fast services and thereby maintain the separating equilibrium while still providing efficiently fast services for the applicants with the highest willingness to pay.

Appendix Table I.A.3 shows that the results from the experiment are inconsistent with this prediction. Column (1) shows the effect on bribes among those who had their applications processed quickly, using 25 working days as the cutoff for if an application was processed quickly. Column (1) shows the effect was positive and that the scorecards increased the bribe payments among

applicants who had their applications processed within 25 working days by BDT 656. Column (2) shows that for applications processed outside of 25 working days limit the estimated effect was a BDT 333 increase but this effect is not statistically significant. Column (3) shows that even for offices that were under-performing at baseline, that had the largest decrease in delays and processing time as a result of the scorecards, the effect of the scorecards on bribes for applications processed within 25 working days is estimated to be an increase of BDT 686. Although the effect is not statistically significantly different from zero, any meaningful negative effect can be rejected.

One potential explanation for the results within the framework of the monopolistic price discrimination model is that the government officials taking the bribes have full information about the applicants' willingness to pay for processing speed. In this information setting the speeding up of processing for those with lower willingness to pay would not affect those with higher willingness to pay, since the bureaucrat could maintain the separating equilibrium by simply requesting different bribes depending on the observed willingness to pay. However, even under full information, the bribe payments are not predicted to increase for those with the highest willingness to pay. The effects of the scorecards on bribes shown in Appendix Table I.A.3 are therefore not consistent with the predictions of the model under any information setting.

Another explanation for why the results are inconsistent with the predictions of the model is that information about the increase in processing speeds had not yet been disseminated to applicants by the time of the survey period. The information treatment is designed to alleviate this problem. Column (4) of Appendix Table I.A.3 shows that, for applications processed within 25 working days, the information treatment had no effect on bribes by itself or in combination with the scorecard treatment. Finally, Column (5) shows that even for applicants that received the information in offices that were under-performing at baseline, no negative effect on bribes can be found for applications processed within 25 working days.

Taken together the results in Appendix Table I.A.3 rejects the model's predictions.

## **B Additional Details on Experiment and Data**

### **B.1 Details on Randomization of Scorecards Treatment Assignment**

The randomization of which land offices were assigned to receive the scorecards was done at the office level. The first wave randomization was done separately for the group of land offices classified by the Government as having full implementation of the e-governance system at the start of the experiment and for the group with partial implementation at the start of the experiment. After these two groups had been separated the randomization strata were created using the following variables:

- Number of applications processed within 45 working days in the months of June and July 2018
- Number of applications pending for more than 45 working days at the end of July 2018

For the group of offices with partial e-governance implementation, stratum were created based on offices being in the first, second or third tertile in the distribution of these variables. Since the group of offices with full e-governance implementation was smaller, stratum were created based on offices being above or below the median in the distribution of these variables.

The second wave randomization was done separately for the group of land offices having received above/below the median number of applications in February and March 2019. After these two groups had been separated the randomization strata were created using the following variables:

- Being in the first, second or third tertile in terms of number of applications processed within 45 working days in March 2019
- Being in the first, second or third tertile in terms of applications pending for more than 45 working days at the end of March 2019

Within each strata, half of the land offices were randomly assigned to treatment.<sup>37</sup> If there was an odd number of offices in a strata, the last office is grouped together with other such "misfits" in their implementation group (first wave) or applications received group (second wave) and half of

---

<sup>37</sup>Random assignment was implemented by the author using the Stata command runiform.

the misfits were randomly assigned treatment. Again, when there were misfits these are grouped together with other misfits from the other implementation group and half of those were assigned treatment. Finally, the last misfits were assigned treatment with a 50% probability.

## **B.2 Data**

### **B.2.1 Administrative Data from E-governance System**

The government partner transferred the data at the beginning of each month from August 2018 until October 2020. Due to privacy concerns the Government only shared administrative data without personal identifying information. The administrative data is at the application level and includes all applications made in the e-governance system since its inception in February 2017. To calculate the number of working days between the date the application started until the end of the application I use data on public, national and general holidays from Time and Date.<sup>38</sup>

Although the performance scorecards are addressed to the current ACL of a land office, the performance is based on how applications made in that land office are processed, regardless of if that ACL was assigned to that office when the application was made or not. I infer what ACL was assigned to what office using the administrative data on what ACL made updates on the applications in a given month. I use administrative user data to separate ACLs from other users and then assign a particular ACL to an office if that ACL is the ACL making the largest number of updates in an office in a particular month. If an office has no updates made by any ACL in a month, I do not assign any ACL to that office in that month, unless there is an ACL that was assigned to that office both prior to and after that month, in which case I conclude that the ACL was assigned to that office without making any updates in that month.

It would be very difficult for any individual bureaucrat to improve their performance scorecard by manipulating the administrative data. The data is stored on a central server that the bureaucrats do not have access to. While it would be possible to create fake applications in the e-governance system, to process these applications with an acceptance the processing fee would have to be paid. Creating fake unprocessed applications would decrease, not increase, a bureaucrat's performance ranking.

---

<sup>38</sup> As of September 2020, available from here: <https://www.timeanddate.com/holidays/bangladesh/>



Some observations were deleted from the administrative data because they were not real applications, but rather applications that had been made for training purposes. During the training of bureaucrats in using the e-governance system, several example applications were made in two land offices that had not yet installed the e-governance system. These "fake" applications were then never removed from the system making it appear as if the e-governance system was active in these two offices. Due to this, these offices were included in the first wave of randomization, one was assigned treatment and one control. In September 2018 I found out that these two offices had not yet installed the e-governance system and I removed all applications from these offices from the administrative data and stopped sending the scorecards to the office that had been assigned the treatment. I also found out that some other applications in the e-governance system are also the result of examples created in training. Using information provided about the dates of the training, I removed applications made before the first wave of randomization suspected to be the result of training. I did not remove any applications made after the start of the experiment.

### **B.2.2 Survey Data**

The survey was carried out in two stages. The average time between the two interviews was 3.3 months. All questions about bribe payments were asked in the second, interview made by phone. Interviewees were given a BDT 50 (USD 0.6) reward in the form of a mobile phone recharge for a completed in-person interview and BDT 100 (USD 1.2) for a completed phone interview.

### **B.2.3 Attrition in the Survey Data**

The overall attrition rate in the survey was 18%. Appendix Table I.A.13 estimate the effect of the scorecard and information treatments on the attrition rate. Column (1) shows that the scorecard treatment is estimated to have had a positive effect on the attrition rate by 3%, an effect which is statistically significant at the 10% level. Columns (2) and (3) show that the information intervention did not affect attrition. Column (4) shows that the effect of the scorecards on attrition is mainly concentrated among offices under-performing at baseline.

If the scorecards caused some applicants to drop out of the study and these applicants, on average, had different values for an outcome variable, this would bias the estimates of the effect

of the scorecards on those outcomes. To assess the potential bias stemming from the differential attrition on the estimated effect of the scorecards on bribe payments, I construct lower Lee bounds for the estimated effect (Lee, 2009). Lower Lee bounds are the relevant robustness check, since the effects on bribe payments are positive (overall and for over-performing offices) or non-negative (for under-performing offices). I create lower Lee bounds by creating a random selection of the applicants from treated offices for whom there is no follow-up survey data, the random selection is equal in size to the estimated effect of the scorecards on the number of applicants not completing the follow-up interview. I then set the bribe payments for this sub-sample to zero, since that is the lowest possible bribe payment. I then conduct the main analysis from Column (2) of Table I.5 and Column (2) of Table I.7. The results are shown in Appendix Table I.A.14. The lower Lee bounds does not qualitatively change the overall results. The estimated effect of the scorecards on bribes is BDT 205. For offices over-performing at baseline the estimated lower Lee bound effect is BDT 639 and still statistically significant. For offices under-performing at baseline the estimated lower Lee bound is BDT -177. The Lee bounds show that even if the entire effect of the scorecards on attrition is on applicants who would have reported zero bribes, this does not substantially change the results.

#### **B.2.4 Cross-validation of Administrative and Survey Data on Processing Times**

There are two application numbers identifying the applications making it possible to match applications from the survey with applications in the administrative data. One is a global identification number automatically generated by the e-governance system. Another is a manually generated local serial number, unique within a year and a land office. The local serial number can also be entered into the e-governance system, but is not a required field and is hence missing for some applications. Unfortunately, few applicants shared their digital application ID, either because they were uncomfortable giving it out or because they could not find the text message through which they received this number. There were many inconsistencies in how the local serial numbers were recorded between and within offices. There also exist several other serial numbers, such as a serial number for the record of rights, that could be confused with the application serial number and that were sometimes reported in the survey instead of the application serial number. This caused

many duplicate serial numbers, even within an office by year combination. Due to these problems, only 45% of the applications from the survey that could be matched to the administrative data. Furthermore, it is possible that incorrect matches between the two data sets were made, although I cannot measure how common this is.

I use these matched applications to rule out some potential, but unlikely, problems with the data. One concern is that bureaucrats receiving the scorecards found a way to manipulate the dates in the administrative data to improve their scorecards. This is unlikely since the administrative data was kept on a government server and could only be accessed by a few government contractors. I worked closely with this group since they were the ones transferring the data to me on a monthly basis and there are no suggestions that they were ever contacted by bureaucrats to alter the data on the server. Another concern is that applicants were pressured into saying that applications were processed faster than they actually were. This is unlikely since the interviews were by phone and there is no way for anyone from the land office to know what the applicant responded. Among the matched applications, the average processing time provided by the applicants in the survey was 67, while the same average time in the administrative data was 89. The average difference between the times was 22 in the control group and 20 in the offices receiving the scorecards. This suggests that there was no differential measurement bias between the treatment and control offices. The correlation between applications being processed within 45 working days in the administrative data and as measured by the processing time stated by the applicants was .25. The correlation in the log of processing times was .24. The low correlations suggests that the matching process suffers from a large number of false positive matches.

Appendix Table I.A.15 shows the results, as estimated in Table I.2, for matched applications only, using both the administrative data and the survey data on processing times. Restricting the applications to only matched observations reduces the precision of the estimates. However, the point estimates for the fraction of applications processed within the 45 working day time limit, the outcome measure most important for the bureaucrats, are similar across the two data sets. This is reassuring since it suggests that the overall effect is not driven by manipulations of the data by the bureaucrats.

## **B.2.5 Comparing Bribe Data with Transparency International Bangladesh Survey Data**

Bribes are notoriously difficult to measure precisely. To validate the magnitude of my bribe estimates I compare them to an independent estimate created by Transparency International Bangladesh as part of their nationally representative National Household Survey (Transparency International Bangladesh, 2016). The survey took place in 2015 and asked about bribes paid between November 2014 and October 2015 and surveyed a nationally representative sample of households. There are two main reasons for why this measure may be different from my estimate of the average bribe payments, other than random differences between samples. First, the survey was done in person, potentially allowing surveyors to build more rapport with the respondents and making them more comfortable to discuss bribe payments. Second, the survey was done in a nationally representative sample of households in Bangladesh and for a different time period. In the Transparency International Bangladesh survey, 605 of the households had made applications for land record changes, among these 57% reported having paid a bribe. Appendix Figure I.A.8 shows that the bribes reported in the Transparency International Bangladesh survey are on average higher than the bribes reported in the phone survey I conducted, but that the difference shrinks substantially when excluding respondents reporting zero bribes. This could be due to that fewer respondents were comfortable to discuss bribe payments over the phone. The typical bribe payments reported in the scorecard experiment survey are slightly higher than the average payment reported in the Transparency International Bangladesh survey but lower than the average non-zero response. Overall, it is reassuring that the two different measures are of similar magnitude, despite using different methodologies, covering different areas, and being done for different time periods.

## **C Additional Empirical Analysis**

### **C.1 External Validity of Results**

The experiment was conducted at a large scale, with more than half of Bangladesh's land offices taking part in the experiment. The large scale of the experiment makes it plausible that the results are externally valid within Bangladesh (Muralidharan and Niehaus, 2017). The experiment included

59% of Bangladesh's land offices, covering an area with a population of approximately 95 million people. The experiment also spans a time frame of 16 months, reducing the concern for novelty effects (Jayaraman et al., 2016), and differences in effects across time periods (Rosenzweig and Udry, 2020). Moreover, the intervention was implemented with the Government of Bangladesh, the same organization that may eventually scale-up the policy. However, the scorecards were designed in a collaboration between a government agency and the author, and produced and distributed by Innovations for Poverty Action, a non-profit research organization. Hence, one should be cautious when extrapolating the results from the experiment to a potential scale-up by the Government itself (Bold et al., 2018). Potential general equilibrium effects within the civil service are discussed in Section 6.3.

### **C.1.1 Geographic External Validity within Bangladesh**

The 311 land offices included in the experiment are the land offices that were actively using the e-governance system by the end of March 2019. Hence, it would be a problem for external validity if offices receiving the scorecard at different times had systematically different effects from the scorecards. In Appendix Table I.A.16, I test if this is the case for the effect on processing times by interacting the treatment with the date the land office started its first application using the e-governance system. Columns (1) and (2) show that the coefficients on the interaction term are close to zero for both the main outcome variables, the fraction of applications processed within the time limit and the overall processing time. In Column (3) the coefficient on the interaction term for the ICW index is -0.005 (S.E. = 0.009). The point estimate is close to zero and my preferred interpretation is that the size of the effect from the scorecards does not vary with the installation date. However, one could interpret the point estimates as a decline in the effect of the scorecards for each month later that the e-governance system was installed. Using a linear prediction, the expected effect of the scorecards for the office in the experiment that had the latest installation date, is expected to be 0.080 standard deviations. The expected effect for the last offices to have the e-governance system installed in Bangladesh, in September 2019, is 0.057 standard deviations. Therefore, although this predicted effect is smaller for offices that had the e-governance system installed later, the scorecards are predicted to have a positive effect for all land offices in Bangladesh

where the e-governance system is installed.

## **C.2 Potential Bias from Applicant Survey and Information Intervention**

Another potential threat to external validity is that the information intervention, or more generally the applicant survey, may have affected the behavior of bureaucrats and applicants. Since the information intervention and the applicant survey were carried out both in offices receiving scorecards and in control offices, it is unlikely that such effects have biased the results. To completely rule out the possibility of such bias, in Appendix Table I.A.17, I conduct the main analysis for processing times from Table I.2 using only applications that could not have been affected by the survey. I restrict the sample to applications from offices that were never surveyed and applications that were made more than 45 working days before the start of the survey in offices that were eventually surveyed. All of the estimates in Appendix Table I.A.17 are very close to the estimates found in Table I.2, ruling out any meaningful bias in the main estimates stemming from interactions between the applicant survey and the scorecards.

## **C.3 Unintended Consequences of the Scorecards**

A common implementation problem of quantitative performance measures is that they lead to unintended and sometimes welfare reducing consequences (Banerjee et al., 2008; Rasul and Rogger, 2018). Below I briefly discuss several potential unintended consequences the performance scorecards could have led to. I do not find evidence for any substantial unintended effect except the effect on bribe payments taken by bureaucrats over-performing at baseline, which I discuss at length in the paper.

### **C.3.1 Improving Indicators without Improving Service Delivery**

One potential concern is that land offices may reduce the number of applications, either by refusing to serve some applicants or by processing some applications using the paper-based system and not fully implement the e-governance system. With a smaller number of applications, it may be easier to reach a higher performance. Anticipating this problem the scorecards measure performance using the absolute number of applications processed within the time limit and not the fraction of

applications. However, the number of applications pending beyond the time limit would still be easier to keep down with fewer applications. Another potential problem could be that bureaucrats only received applications which they knew would be easier to process. The size of the land for which the land record change is being made is positively associated with the processing time. Therefore, if bureaucrats intended to avoid accepting complex applications we would also see a decrease in the average land size.

Appendix Table I.A.18 Column (1) shows that the scorecards did not substantially affect the number of applications received in the e-governance system. Column (2) shows that the scorecards did not decrease the number of scorecards more in the offices that were under-performing at baseline and where the scorecards have the largest effect on processing times. Column (3) shows that the scorecards did not substantially affect the average land size among applications received, and Column (4) shows no evidence of a heterogeneous effect on land size among received applications. Overall the results are consistent with the bureaucrats not altering the applications received in response to the scorecards.

### **C.3.2 Quality of Decision Making**

Another potential concern is that the quality of the decisions made by the bureaucrats was reduced by the scorecards. The main decision the bureaucrat makes with regards to the application is whether to reject or accept it. It is possible that when the bureaucrat spends less time on each application, more acceptances or more rejections are made depending on what the quickest action is to dispose of the application. Column (1) of Appendix Table I.A.19 shows that the scorecards did not change the fraction of applications accepted. However, it is still possible that the quality of the decision was worse, meaning that more applications that should have been rejected were accepted and that more applications that should have been rejected were accepted. If an application is wrongfully rejected, applicants typically reapply in the same office. Therefore, the fraction of applicants reapplying after having been previously rejected can be used as an indicator for the fraction of incorrect rejections. Column (2) of Appendix Table I.A.19 shows that the fraction of applicants stating that they were reapplying, after having been previously rejected for the same application, was not substantially increased by the scorecards. Together, these results suggest that

the scorecards did not lead to a decrease in the quality of decision making.

#### **C.4 Effects of Information Intervention on Bribes**

Appendix Table I.A.4 shows that the information treatment did not have an effect on bribes, neither by itself nor in combination with the scorecards. Column (1) and (4) show that the information treatment by itself did not have a substantial impact on reported bribes or estimates of typical bribe payments. Columns (2) and (5) show that even together with the information treatment, the scorecards did not reduce bribes, neither for reported bribes nor for typical bribes. Columns (3) and (6) show that even among under-performing land offices, that improved the processing times the most, bribes did not decrease, neither with nor without the information treatment. Columns (3) and (6) also show that the positive effect of the scorecards on bribes, among offices over-performing at baseline, is similar across for applicants receiving the information intervention and applicants not receiving the information intervention.

#### **C.5 Effects of Scorecard on Applicant Satisfaction**

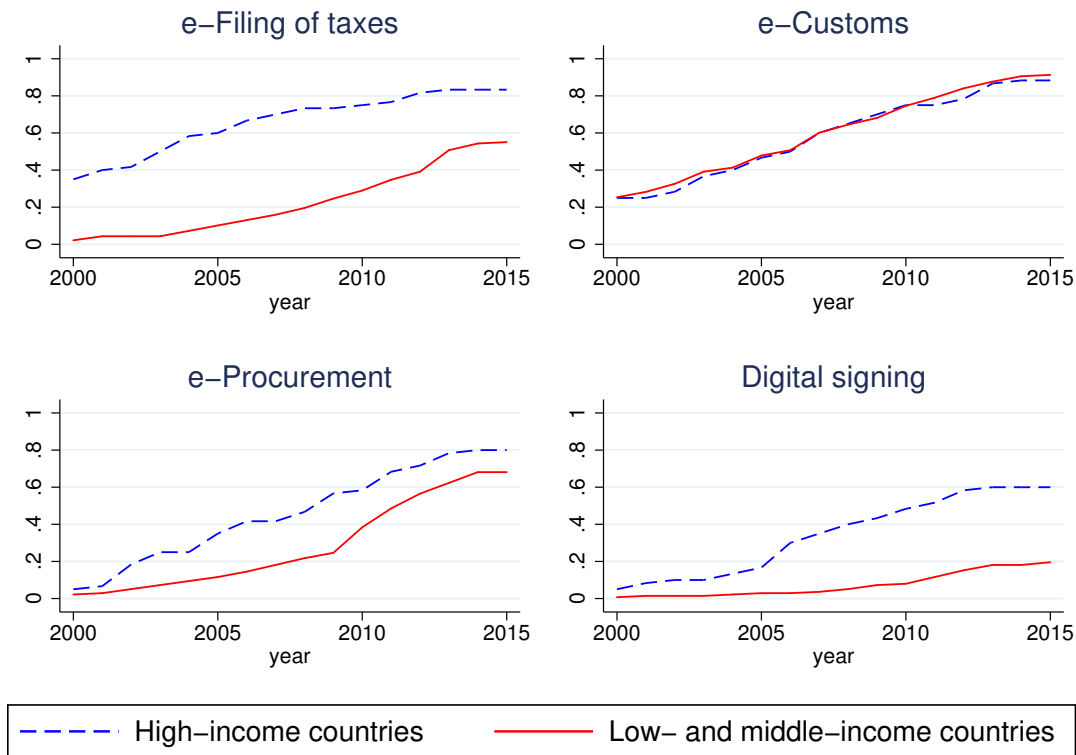
Appendix Table I.A.20 shows the estimated effects of the scorecards on applicant satisfaction. Satisfaction was measured in the follow-up phone survey by asking applicants "Overall, how satisfied are you with the processing of your application?". The respondent could answer the question on a five-point scale ranging from very satisfied to not satisfied at all. The response was then transformed into standard deviations from the control group mean and used as an outcome variables in regression Equation 1 and 2. Column (1) of Appendix Table I.A.20 shows the overall effect on satisfaction which is negative but small and not statistically significant. Column (2) splits up this effect between offices that were under-performing and offices over-performing at baseline. The negative effect is driven by offices that were over-performing at baseline which is consistent with the observation that the scorecards increased bribe payments increased in these offices. Furthermore, despite that the scorecards were successful in reducing processing for offices under-performing at baseline, the effect on the satisfaction stated by applicants in these offices is close to zero. Overall the results are consistent with a low valuation of faster processing times by the applicants but given the imprecise results it is hard to draw any definitive conclusion from the



null result.

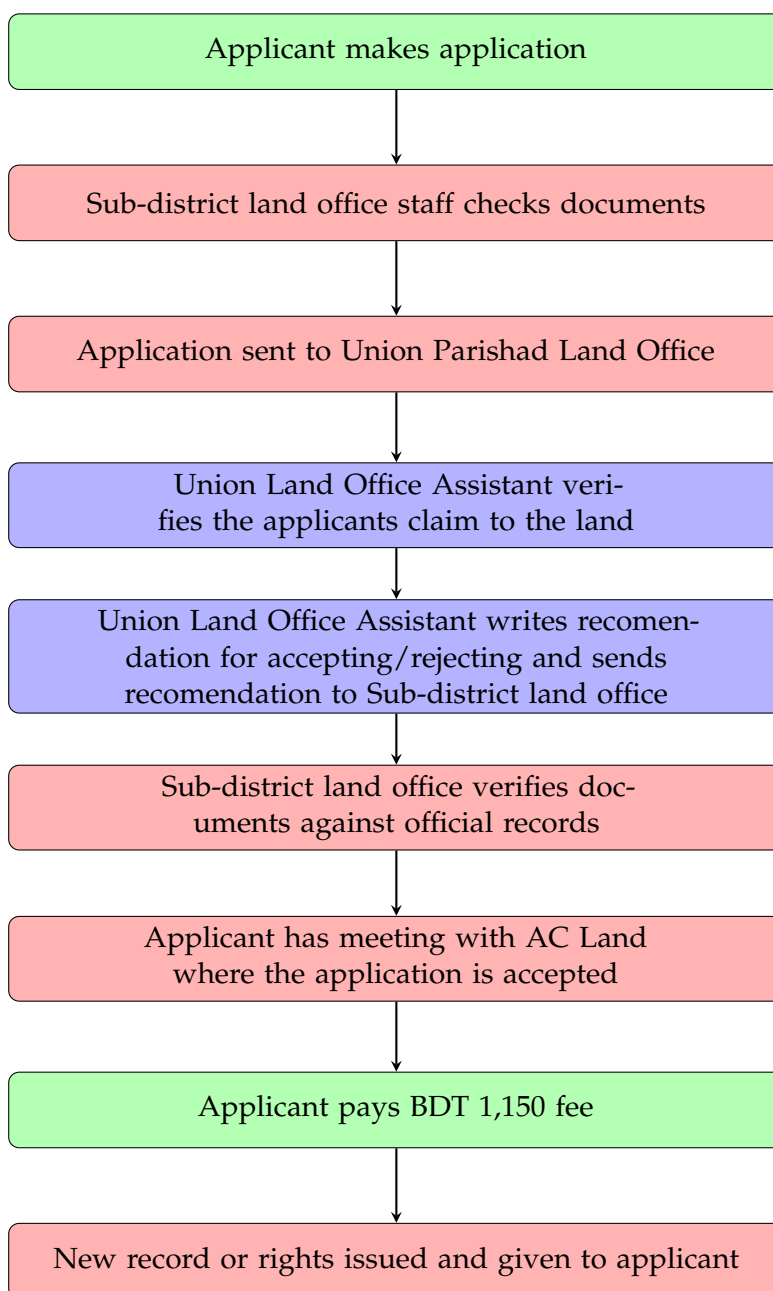
## **D Additional Tables and Figures**

Figure I.A.1: Digital Government Capacity by Income Group



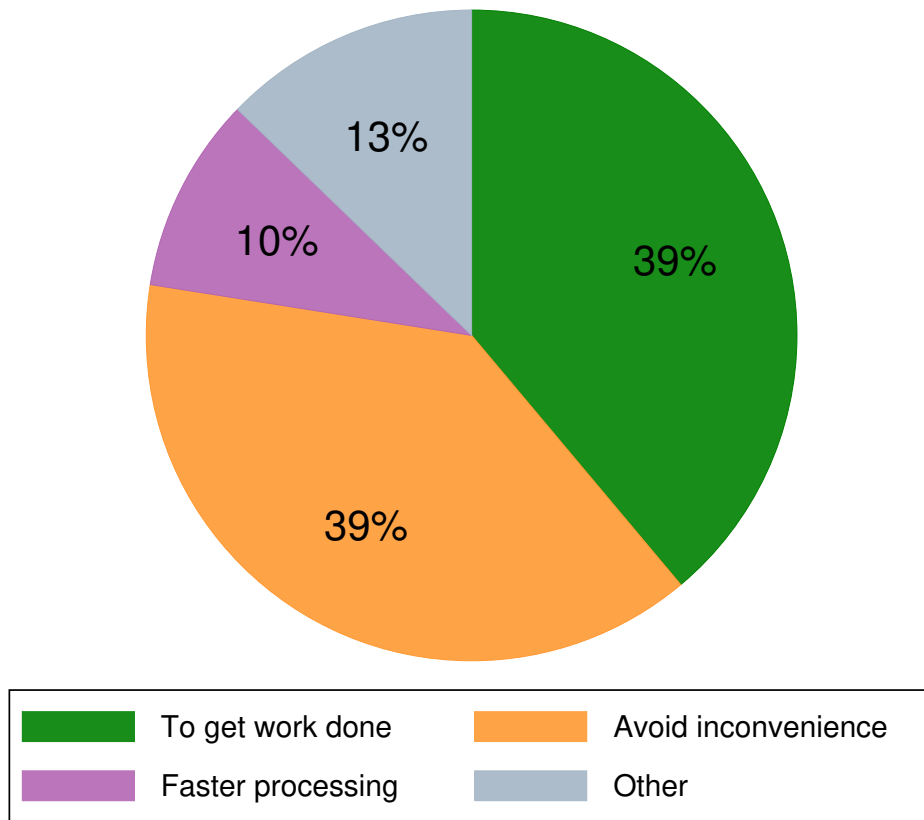
This figure shows what fraction of countries have at least a partially implemented e-governance system for four common interactions between the government and citizens or firms. The first is an e-governance system for filing taxes. The second is an e-governance system for clearing customs. The third is an e-governance system for public procurement. The fourth is legislation enabling digital signatures. Data is from the World Bank's *Public Financial Management Systems And E-Services Global Dataset* updated August 2017. Discussed in Section 2.

Figure I.A.2: Application Process for Successful Application



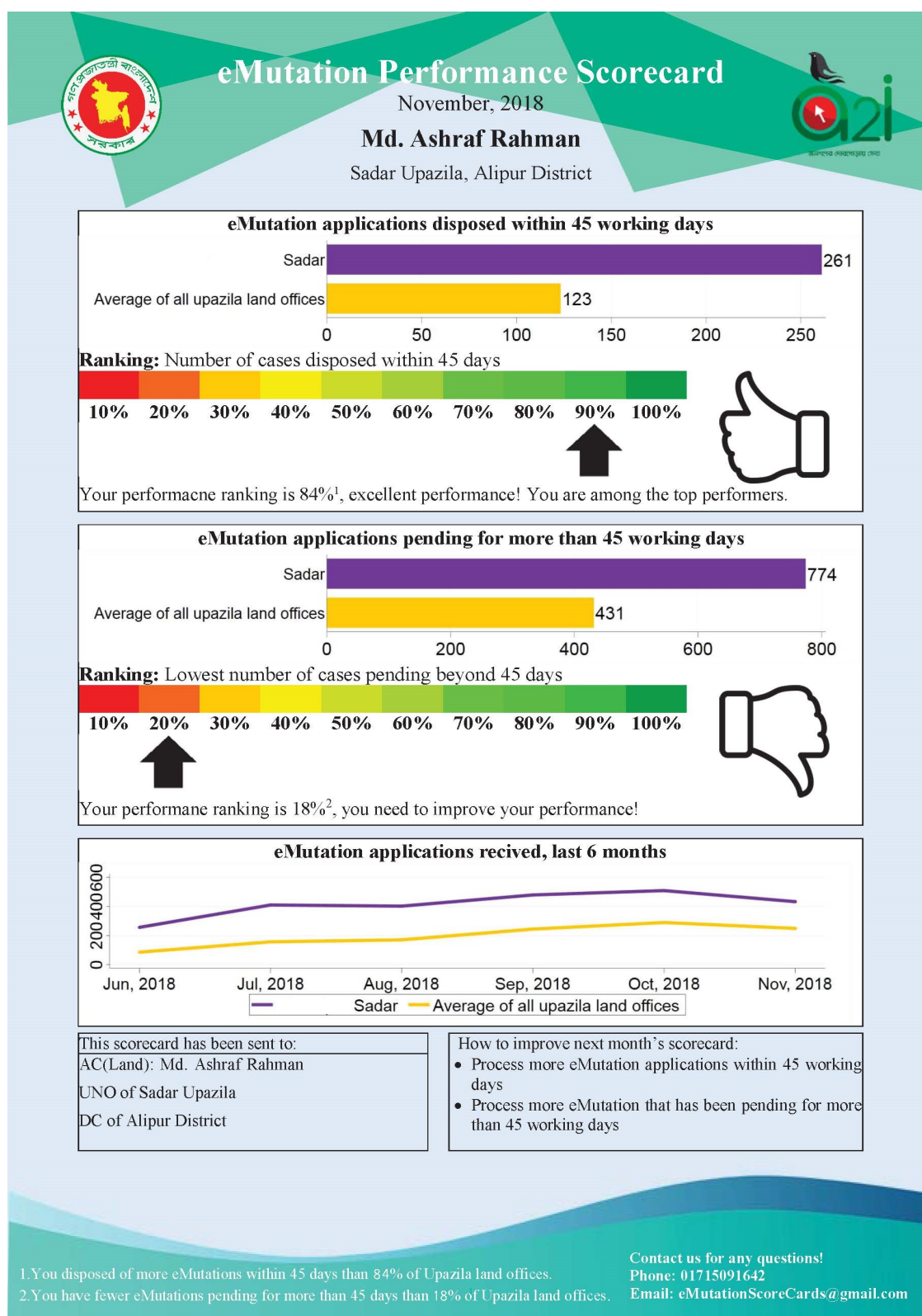
This figure provides a visual overview of the process of getting an application for a land record change approved. Green boxes represents actions by the applicant. Red boxes represents actions by the sub-district (*Upazila*) land office. Blue boxes represents actions by the local (*Union Parishad*) land office.

Figure I.A.3: Stated Reasons for Bribe Payments



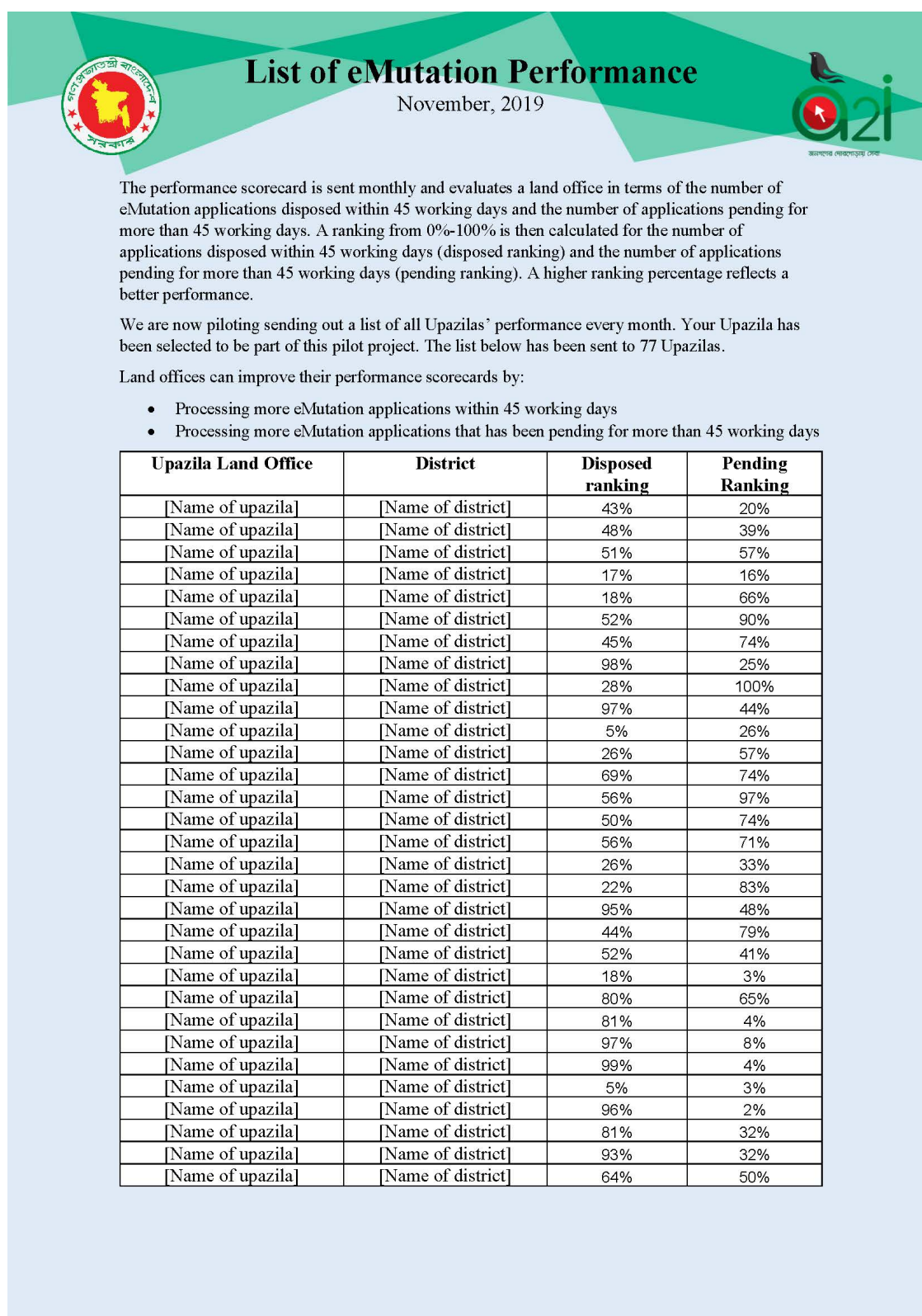
This figure shows the reason stated by the applicants for paying bribes. The responses are weighted by the amount of the bribe. Therefore, the percentages should be interpreted as what percentage of the total bribe amounts were paid for what reason. The question was open-ended and was coded into response categories. Discussed in Sections 2.1 and 6.

Figure I.A.4: Example of Performance Scorecard



This is an example of a performance scorecard in English. The ACL name and land office name are changed to preserve the anonymity of the civil servant. A version in Bengali was also included.

Figure I.A.5: Example Peer Performance List

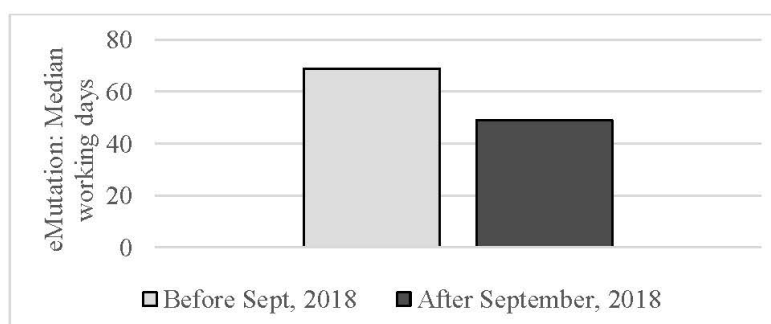


This is an example of the first page of a list of a peer performance list, the full list contain two pages. The office and district names have been removed to preserve the anonymity of the civil servants.

Figure I.A.6: Information Pamphlet Given in Information Intervention

## Information for applicants Land Record eMutation

Over the past 6 months the Government of Bangladesh have taken several steps to reduce the time it takes to process a Land Record Mutation. Before a typical Land Record eMutation took 69 working days (more than 3 months) now a typical Land Record eMutation takes 49 working days (a little more than 2 months).



You can apply for a Land Record eMutation by visiting the Upazila Land Office or from any computer connected to the internet (<http://training.land.gov.bd/mutation/application>).

Steps of Land Record eMutation and timeline:

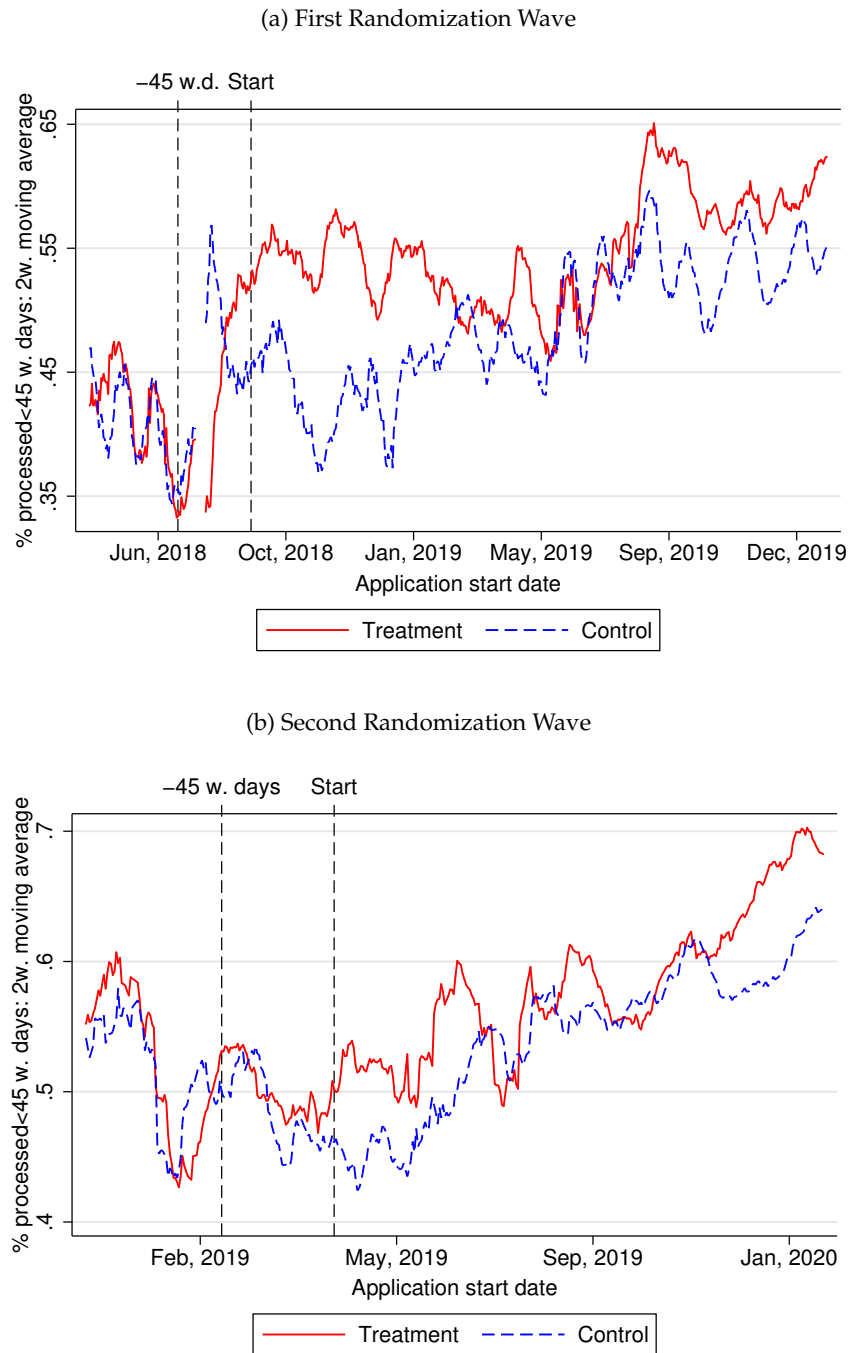
1. Make application online or in Upazila Land Office
2. Upazila Land Office will check the application and send it to Union Land Office
3. Union Land Office Assistant will make visit to land and write report to Upazila Land Office
4. Upazila Land Office will read report and call you for hearing via text message
5. You will attend hearing (according to text message)
6. Pay fee of 1150 taka and receive your Khatian

This information sheet was prepared by Innovations for Poverty Action in collaboration with a2i and the Land Reforms Board of Bangladesh.

Contact phone number: [REDACTED]

English translation of the information pamphlet given to applicants in the information intervention. The pamphlet shows the median application time for processed applications made before September 2018 and applications made after September 2018 in the 112 offices where the interviews took place. The data is as of February 2019. The same pamphlet was given to applicants in both treatment and control offices.

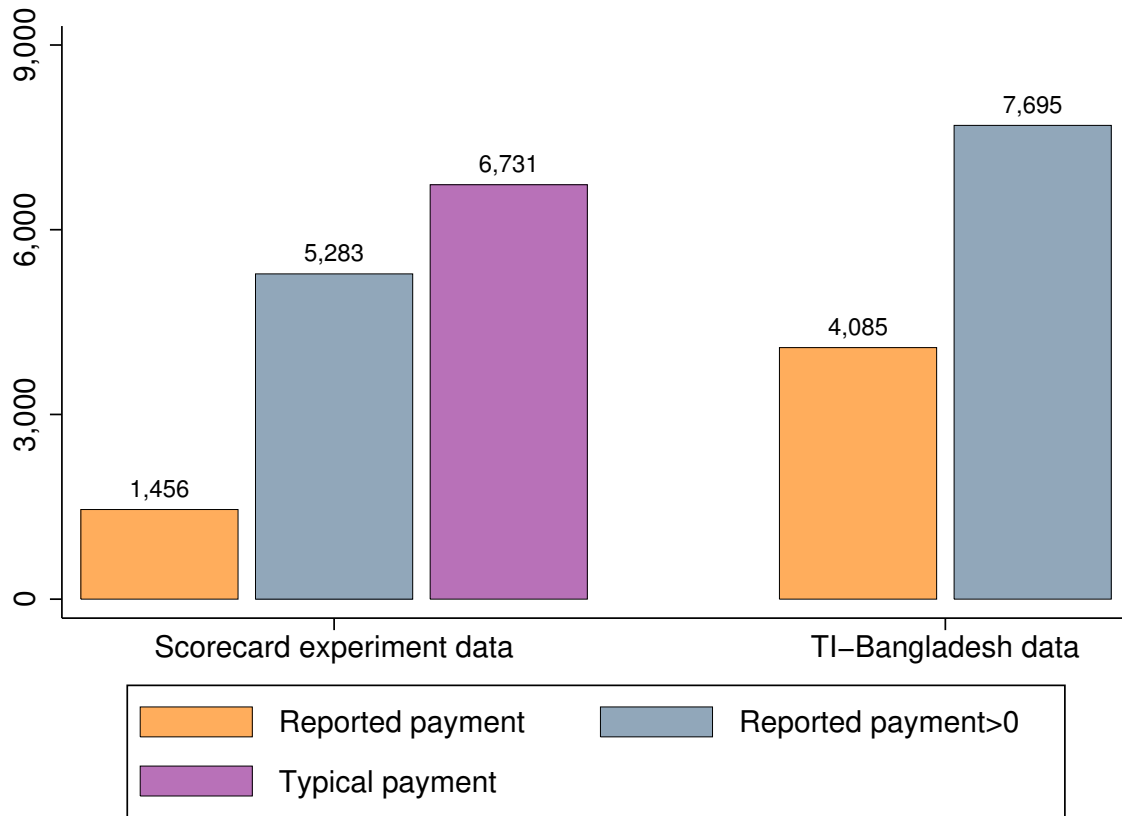
Figure I.A.7: Fraction of Applications Processed within 45 Working Days



This figure shows the two-week moving averages of the fraction of applications processed within the 45 working day limit in the treatment and control groups. Sub-figure I.A.8a shows data from the first randomization wave Sub-figure I.A.8b shows data from the second randomization wave. The effect of the treatment can in principle have started for applications started 45 working days before the first scorecard was sent (first vertical dashed line) but only application made after the first scorecard was sent (second vertical dashed line) were fully treated. The gap in the Sub-Figure I.A.8a time-line is due to a server error that caused the e-governance system to temporarily shut down in late July 2018. Discussed in Section 4.1.1.



Figure I.A.8: Comparison of Estimated Bribes to Transparency International Bangladesh Survey



This figure shows the average bribe payments reported in the phone survey conducted to evaluate the scorecard experiment and in an independent survey by Transparency International Bangladesh. The first bar shows the average value of bribe payments reported by the applicant in the scorecard experiment phone survey, 73% of the applicants reported having paid no bribes. The second bar shows the average value of bribe payments reported by applicants reporting having paid some bribe in the scorecard experiment phone survey. The third bar shows the average value of an estimated "typical bribe payment by a person like yourself" reported in the scorecard experiment phone survey, 27% of the applicants responding to this question reported that a typical applicant paid no bribes. The fourth bar shows the average value of bribe payments reported by the applicants in the Transparency International Bangladesh survey, 57% of the respondents reported having paid no bribe for their land record change. The fifth bar shows the average value of bribe payments reported by respondents reporting having paid some bribe in the Transparency International Bangladesh survey. All variables are winsorized at the 99th percentile. Observations in the three first bars are inversely weighted by the number of observations in that land office. Discussed in Appendix Section B.2.5.

Table I.A.1: Balance of Randomization: Administrative Data

Variable	(1) Control		(2) Scorecard		(3) T-test (1)-(2)
	N/[Clusters]	Mean/SD	N/[Clusters]	Mean/SD	Difference/SE
<45 w. days	56,686 [146]	0.422 (0.49)	57,028 [146]	0.444 (0.50)	-0.022 (0.08)
IHS(w. days)	56,686 [146]	4.744 (1.02)	57,028 [146]	4.734 (1.11)	0.010 (0.199)
Process time (w. days)	56,686 [146]	88.274 (78.75)	57,028 [146]	96.456 (95.96)	-8.181 (20.2)
Approved before experiment start	38,728 [141]	0.738 (0.44)	36,272 [136]	0.751 (0.43)	-0.013 (0.06)

This table shows the balance of randomization for treatment and control offices using administrative data from 45 working days before the first scorecard was sent (this date is different for randomization wave 1 and randomization wave 2 offices). Due to this restriction only 292 of the 311 offices are part of the balance of randomization data. Applications not processed by the first scorecard had the processing time imputed using the procedure described in Section 2.5.1. Data on approvals are as per the start of the treatment. P-value for F-test of joint orthogonality: 0.83. Observations are uniformly weighted. When using the weighted regression specification from Equation 1 on this data the effect of the treatment is not statistically significantly different from zero for any of the outcome variables. \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ . Discussed in Section 2.6.

Table I.A.2: Balance of Randomization: Survey Data

Variable	(1) Control		(2) Scorecard		(3) T-test (1)-(2)
	N/[Clusters]	Mean/SD	N/[Clusters]	Mean/SD	Difference/SE
Applicant age	1,463 [56]	47.33 (13.83)	1,440 [56]	47.37 (13.20)	-0.04 (0.62)
Female	1,498 [56]	0.07 (0.25)	1,520 [56]	0.06 (0.24)	0.01 (0.01)
Monthly income (BDT)	1,407 [56]	28,505 (92,587)	1,384 [56]	32,568 (133,391)	-4,063 (5,832)
App. status: Applying	1,498 [56]	0.24 (0.42)	1,520 [56]	0.20 (0.40)	0.03 (0.04)
Ongoing	1,498 [56]	0.60 (0.49)	1,520 [56]	0.61 (0.49)	-0.004 (0.04)
Rejected	1,498 [56]	0.002 (0.04)	1,520 [56]	0.005 (0.07)	-0.003 (0.003)
Approved	1,498 [56]	0.07 (0.26)	1,520 [56]	0.07 (0.25)	0.01 (0.02)
Land Value (BDT 100k)	1,418 [56]	18.21 (44.30)	1,382 [56]	22.09 (53.42)	-3.87 (2.85)
Land Size (Decimal)	1,455 [56]	25.42 (43.44)	1,437 [56]	25.66 (52.61)	-0.24 (2.91)

All data comes from the in-person survey of applicants, which was conducted before the conclusion of the processing of the application, but after the start of the scorecards. USD/BDT $\approx$ 84.3. 1 decimal = 1/100 acre. P-value for F-test of joint orthogonality: 0.89. Observations are uniformly weighted. When using the weighted regression specification from Equation 1 on this data, the effect of the scorecards is not significant at the 5% for any of the outcome variables, and significant at the 10% level only for land value. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section 2.6.

Table I.A.3: Testing Prediction from Monopolistic Price Discrimination Model

	Reported payment				
	(1)	(2)	(3)	(4)	(5)
Scorecard	656.4*** (214.6)	332.7 (259.8)		903.3*** (319.2)	
Scorecard x Overperform			635.2*** (227.2)		
Scorecard x Underperform			686.2* (399.1)		
Overperform baseline			9.139 (322.2)		12.19 (317.6)
Information treatment				-111.2 (239.1)	-110.0 (238.2)
Scorecard x Information				-430.4 (431.3)	
Info x Scorecard x Underperform					499.7 (512.0)
No info x Scorecard x Underperform					946.6* (495.2)
Info x Scorecard x Overperform					452.2 (295.5)
No info x Scorecard x Overperform					875.5** (385.7)
Start month FE	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes
Sample	Appr. < 25	Appr. > 25	Appr. < 25	Appr. < 25	Appr. < 25
Observations	672	1,447	672	672	672
Clusters	109	111	109	109	109

This table shows the effect of the scorecard and information treatments on bribes made for application processing. Column (1) and (3)-(5) use data only from applications that were approved within 25 working days while Column (2) uses data from applications that have an approval time of more than 25 working days. USD/BDT≈84.3 Continuous variables winsorized at the 99th percentile. Standard errors are clustered at the land office level. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Appendix Section A.2.

Table I.A.4: Effect of Information Treatment and Scorecards on Bribes

	Reported payment			Typical payment		
	(1)	(2)	(3)	(4)	(5)	(6)
Information treatment	1.638 (146.5)	-10.88 (187.0)		252.1 (495.4)	-189.9 (749.7)	
Scorecard		261.0 (222.6)			644.9 (720.1)	
Scorecard x Information		8.499 (274.0)			812.3 (1098.8)	
Info x Scorecard x Overperform			674.2** (268.4)			1761.7** (871.3)
No info x Scorecard x Overperform			603.9** (253.3)			2747.7*** (873.7)
Info x Scorecard x Underperform			-94.80 (287.0)			999.7 (1239.5)
No info x Scorecard x Underperform			-17.53 (315.7)			-1440.1 (875.8)
Overperform baseline			-819.9*** (287.9)			-1766.7* (924.2)
Start month FE	Yes	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,018	3,018	3,018	1,896	1,896	1,896
Clusters	570	112	112	544	112	112
Control mean	1,467	1,302		6,512	6,157	

This table shows the effect of the scorecard and information treatments on bribes made for application processing. Column (1)-(3) estimate the effects on reported bribes. Columns (4)-(6) estimate the effects on typical bribe payments. USD/BDT $\approx$ 84.3 All outcome variables are winsorized at the 99th percentile. In Columns (1) and (4) standard errors are clustered at the land office by day level. In Columns (2)-(3) and (5)-(6) standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Appendix Section 5.1.

Table I.A.5: Robustness to Imputation Technique: Effect on Processing Times

	(1) Mean	(2) Office mean	(3) Rand. obs.	(4) No impute
Scorecard	-0.125** (0.0593)	-0.119** (0.0579)	-0.123** (0.0582)	-0.120** (0.0511)
Constant	4.417*** (0.0403)	4.402*** (0.0389)	4.406*** (0.0395)	4.294*** (0.0353)
Start month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	1,050,924	1,050,924	1,050,924	972,589
Clusters	311	311	311	311

This table shows the robustness of the result in Column (2) of Table I.2 with regards to the imputation procedure used to assign a processing time to applications that are not yet processed. Column (1) uses the mean of processing times for all applications that were processed after the number of days that the application that I am imputing the processing time for has been pending. Column (2) uses the mean of processing times for applications in that land office that were processed after the number of days that the application that I am imputing the processing time for has been pending. Column (3) uses a randomly selected processing time of the processing times that are larger than the number of days that the application that I am imputing the processing time for has been pending. Column (4) drops all applications that are not yet processed from the sample. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ . Discussed in Section 4.1.

Table I.A.6: Robustness to Functional Form Assumption: Effect on Processing Times

	ln(Days)	Working Days		
	(1)	(2)	(3)	(4)
Scorecard	-0.123** (0.0587)	-6.492 (4.033)	-0.101 (0.0646)	-0.0980* (0.0585)
Start month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	1,048,876	1,050,924	1,050,924	1,050,924
Clusters	311	311	311	311
Specification	OLS	OLS	Poisson	Neg. Binomial

This table shows the robustness of the effect of the scorecards on the processing time to different assumptions regarding the functional form of the relationship between the processing time and the scorecards. Column (1) uses the natural logarithm transformation treating observations with the value zero as missing. Column (2) uses the untransformed number of working days. Column (3) shows the results of a Poisson regression. Column (4) shows the results of a negative binomial regression. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ . Discussed in Section 4.1.

Table I.A.7: Robustness to Alternative Specifications: Effect on Processing Times

	ICW Index				
	(1)	(2)	(3)	(4)	(5)
Panel A. Overall effect					
Scorecard	0.081 (0.091)	0.125* (0.069)	0.126** (0.059)	0.091 (0.078)	0.129** (0.055)
Panel B. Heterogeneous effects					
Scorecard x Overperform baseline	-0.016 (0.115)	-0.002 (0.089)	0.002 (0.078)	0.023 (0.097)	0.020 (0.075)
Scorecard x Underperform baseline	0.210* (0.121)	0.283*** (0.093)	0.272*** (0.087)	0.190* (0.113)	0.259*** (0.082)
Overperform baseline	0.558*** (0.108)	0.580*** (0.087)	0.311*** (0.090)	0.497*** (0.131)	0.237** (0.107)
Start month FE	No	No	No	Yes	Yes
Stratum FE	No	No	No	Yes	Yes
Weighted by office	No	Yes	Yes	No	Yes
Baseline controls	No	No	Yes	No	Yes
Observations	1,050,924	1,050,924	1,050,924	1,050,924	1,050,924
Clusters	311	311	311	311	311

This table shows the robustness of the estimated effect of the scorecards on the IWC Index of being processed within the time limit and the IHS of processing time to different regression specifications. Panel A shows the estimates of the overall effect similar to the estimates in Table I.2. Panel B shows the estimates of the heterogeneous effects, similar to the estimates in Table I.6. In what follows I describe how the specifications differ from the specifications in those tables. Column (1) shows the estimate from an unweighted regression with no fixed effects. Column (2) shows the estimate from a regression with no fixed effects. Column (3) shows the estimate from a regression with no fixed effects, controlling for baseline measures of the number of applications processed within 45 working days, the number of applications pending beyond 45 working days, applications received and the fraction of applications processed within 45 working days. Column (4) shows the estimate from an unweighted regression. Column (5) shows the estimate from a regression controlling for baseline measures of the number of applications processed within 45 working days, the number of applications pending beyond 45 working days, applications received and the fraction of applications processed within 45 working days. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section 4.1.



Table I.A.8: Effect on Office by Month Level Outcomes

Panel A. Overall effect	(1)	(2)	(3)	(4)	(5)
	IHS Dis. ≤ 45	IHS Pen. > 45	Rank dis.	Rank pen.	ICW index
Scorecard	0.212* (0.121)	-0.058 (0.140)	2.133 (1.685)	1.464 (1.782)	0.080 (0.062)
Panel B. Heterogeneous effects					
Scorecard x Overperform	-0.039 (0.149)	0.280 (0.212)	-0.819 (2.209)	-2.524 (2.609)	-0.099 (0.082)
Scorecard x Underperform baseline	0.499** (0.195)	-0.441** (0.180)	5.563** (2.587)	5.899** (2.356)	0.285*** (0.091)
1st PS>median	0.537** (0.236)	-0.286 (0.264)	7.685** (3.425)	1.944 (3.401)	0.242* (0.127)
Month FE	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes
Observations	4,516	4,516	4,516	4,516	4,516
Clusters	311	311	311	311	311

This table shows the effect of the scorecards on office by month level outcomes. Panel A shows the estimates of the overall effect similar to the estimates in Table I.2. Panel B shows the estimates of the heterogeneous effects, similar to the estimates in Table I.6. Column (1) shows the effect on the IHS of the number of applications processed within 45 working days. Column (2) shows the effect on the IHS of the number of applications pending beyond 45 working days. Column (3) shows the effect on the percentile ranking in terms of the number of applications processed within 45 working days. Column (4) shows the effect on the percentile ranking in terms of the number of applications processed within 45 working days, a higher number of pending applications leads to a lower ranking. Column (5) shows the result on a ICW index created with the outcome variables of Tables (1)-(4). Standard errors are clustered at the land office level. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section 4.1.

Table I.A.9: Robustness to Alternative Specifications: Effect on Bribes

	Typical payments					Reported payments					To gov. off.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. Overall effect												
Scorecard	1,324* (699)	1,277* (700)	978 (613)	1,974** (844)	566 (443)	384* (201)	355* (195)	288 (174)	564** (263)	175 (125)	214 (137)	
Panel B. Heterogeneous effects												
Scorecard x Overperform	2,294** (931)	2,034** (877)	2,467*** (845)	3,055*** (1,070)	1,637*** (583)	741*** (235)	708*** (215)	680*** (229)	1,066** (426)	454*** (163)	416** (165)	
Scorecard x Underperform	435 (1,011)	557 (1,062)	-486 (883)	938 (1,309)	-480 (640)	67 (312)	29 (307)	-52 (245)	80 (364)	-66 (168)	55 (197)	
Overperform baseline	-1,314 (813)	-1,504 (916)	-2,010** (921)	-1,945* (1,100)	-1,616** (688)	-515* (262)	-700*** (266)	-778*** (282)	-660* (386)	-637*** (195)	-585** (227)	
Start month FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	
Stratum FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	
Weighted by office	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	
Winsorized	99 pctl.	99 pctl.	99 pctl.	No	95 pctl.	99 pctl.	99 pctl.	99 pctl.	No	95 pctl.	99 pctl.	
Observations	1,896	1,896	1,896	1,896	1,896	3,018	3,018	3,018	3,018	3,018	3,018	
Clusters	112	112	112	112	112	112	112	112	112	112	112	

This table shows the robustness of the estimated effect of the scorecards on bribes. Panel A shows the estimates of the overall effect similar to the estimates in Columns (1) and (2) of Table I.5. Panel B shows the estimates of the heterogeneous effects, similar to the estimates in Columns (1) and (2) of Table I.7. Columns (1)-(5) show the effect on the estimate for how much a "normal person, like yourself" pays in bribes. Columns (6)-(10) show the effect on reported payments to government officials or agents beyond the official fee. In what follows I describe how the specifications differ from the specifications in Tables I.5 and I.7. Columns (1) and (6) show the estimate from an unweighted regression with no fixed effects. Columns (2) and (7) show the estimate from a regression with no fixed effects. Columns (3) and (8) show the estimate from an unweighted regression. Columns (4) and (9) show the estimate when not winsorizing the outcome variable. Columns (5) and (10) show the estimate when winsorizing the outcome variable at the 95th percentile. Column (11) shows the estimate of the effect on payments made directly to government officials, excluding payments made to agents. Standard errors are clustered at the land office level. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section 4.4.

Table I.A.10: Robustness to Measures of Baseline Performance: Effect Heterogeneity

	Processing time: ICW Index			Typical payments			Reported payments		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Scorecard x Overperform 3m baseline	0.0157 (0.0802)			2097.1*** (712.1)			618.6*** (213.8)		
Scorecard x Underperform 3m baseline	0.261*** (0.0867)			125.9 (940.1)			-28.53 (258.4)		
Treat x 4th quartile baseline		-0.0343 (0.104)			2335.1** (923.3)			550.2** (271.2)	
Treat x 3rd quartile baseline		0.0186 (0.119)			2425.3** (1182.4)			758.4** (365.2)	
Treat x 2nd quartile baseline		0.172 (0.122)			0.156 (967.7)			111.6 (299.4)	
Treat x 1st quartile baseline		0.353*** (0.118)			-234.2 (1510.1)			-186.7 (396.0)	
Treat x Baseline ranking			-0.00731** (0.00305)			41.23 (29.05)			16.07* (8.840)
Scorecard			0.116** (0.0573)			1090.6* (602.5)			299.8* (171.4)
P-value sub-group diff.	0.04			0.10			0.05		
Start month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline performance control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,050,924	1,050,924	1,050,924	1,896	1,896	1,896	3,018	3,018	3,018
Clusters	311	311	311	112	112	112	112	112	112

This table shows the robustness of the results for the heterogeneity of the effect of the scorecards for different baseline performance measures. Columns (1)-(3) show the effects on the index of the two main processing time outcome variables used in Column (3) of Table I.2. Columns (4)-(6) show the effect on the estimate for how much a "normal person, like yourself" pays in bribes. Columns (7)-(9) show the effects on the bribe payments reported by the applicant. Columns (1), (4), and (7) show the heterogeneity in the effect of the scorecards based on the office having an above- or below-median average ranking across the last three months of the baseline period. Columns (2), (5), and (8) show the heterogeneity based on the quartile of baseline ranking. Columns (3), (6), and (9) show the heterogeneity based on the continuous baseline ranking. USD/BDT≈84.3. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section 4.5.

Table I.A.11: Effects on Expected Processing Time

	ln(Expected processing time)				
	(1)	(2)	(3)	(4)	(5)
Scorecard	-0.0886** (0.0439)		-0.0778 (0.0511)		
Information treatment		-0.0373 (0.0276)	-0.0262 (0.0235)		
Scorecard x Information			-0.0211 (0.0439)		
Scorecard x Overperform baseline				-0.0640 (0.0682)	
Scorecard x Underperform baseline				-0.0944* (0.0562)	
Info x Scorecard x Overperform					-0.0758 (0.0666)
No info x Scorecard x Overperform					-0.0526 (0.0784)
Info x Scorecard x Underperform					-0.133* (0.0681)
No info x Scorecard x Underperform					-0.0526 (0.0549)
Overperform baseline				-0.137* (0.0691)	-0.138** (0.0690)
Start month FE	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes
Observations	2,657	2,657	2,657	2,657	2,657
Clusters	112	528	112	112	112
Control mean	56	57	57		

This table shows the effect of the scorecard and information treatments on expected processing times at the time of the in-person interview. The outcomes variable is winsorized at the 99th percentile. In Column (1) standard errors are clustered at the land office by day level. In Columns (2) and (3) standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section 6.4.

Table I.A.12: Effect on Bureaucrat Transfers

	(1)	(2)	(3)	(4)	(5)	(6)
	Transfer	Transfer	Duration	Duration	No ACL	No ACL
Scorecard	0.00210 (0.00554)		0.533 (0.671)		-0.000236 (0.0241)	
Scorecard x Overperform		0.00135 (0.00792)		0.503 (0.891)		0.0103 (0.0335)
Scorecard x Underperform		0.00288 (0.00821)		0.563 (1.037)		-0.0125 (0.0364)
Overperform baseline		0.00589 (0.00962)		-0.202 (1.135)		-0.00531 (0.0346)
P-value: sub-group diff.		0.90		0.97		0.65
Month FE	Yes	Yes	No	No	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,516	4,516	304	304	4,516	4,516
Clusters	311	311	304	304	311	311
Control mean	0.07		12.22		0.13	
Control mean		0.07		12.19		0.12
Overperformers: control mean		0.07		12.26		0.14

This table shows the effect of the scorecards on transfers of ACLs. Columns (1) and (2) show the effects on the fraction of ACLs transferred away from the office in a particular office-month, using data for each office month after the start of the experiment until the last month of the experiment (March 2020). Columns (3) and (4) show the effect on the duration of the posting in months for the first bureaucrat to hold the position as ACL in each of the offices in the experiment. Columns (5) and (6) show the effect on not having any ACL in a particular office-month. The data is administrative data from the e-governance system. Standard errors are clustered at the land office level, except for in Columns (3) and (4), where heteroskedasticity robust standard errors are used. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section 6.4.2.

Table I.A.13: Treatment Effects on Survey Attrition

	Attrition			
	(1)	(2)	(3)	(4)
Scorecard treatment	0.0293*		0.0256	
	(0.0160)		(0.0194)	
Information treatment		0.00223	-0.00254	
		(0.0121)	(0.0174)	
Scorecard x Information			0.00708	
			(0.0233)	
Scorecard x Overperform baseline				0.0123
				(0.0214)
Scorecard x Underperform baseline				0.0476**
				(0.0240)
Overperform baseline				0.00461
				(0.0253)
Start month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	3,696	3,695	3,695	3,696
Clusters	112	112	112	112

This table shows the effect of the scorecards and information treatment on attrition from the survey. Attrition is measured from a survey being attempted at the time of the in-person survey until having a successful follow-up survey by phone. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Section B.2.2.

Table I.A.14: Lower Lee Bounds for Effects on Bribes

	Reported payment	
	(1)	(2)
Scorecard treatment	204.88 (179.72)	
Scorecard x Overperform baseline		638.51*** (224.25)
Scorecard x Underperform baseline		-177.14 (255.68)
Overperform baseline		-820.76*** (287)
Start month FE	Yes	Yes
Stratum FE	Yes	Yes
Weighted by office	Yes	Yes
Observations	3,075	3,076
Clusters	112	112
Control mean	1,278	
Overperformers: Control mean		916
Underperformers: Control mean		1,616

This table shows the lower Lee bounds (Lee, 2009) for the estimates of the effects of the scorecards on bribe payments shown in Column (1) of Table I.5 and Column (1) of Table I.7. A number, equal to the differential attrition rate, of randomly selected observations from the treatment group are added back into the data and assigned a value of zero for bribe payments. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ . The table is discussed in Appendix Section B.2.3.

Table I.A.15: Comparison of Effects in Administrative and Survey Data

	(1)	(2)	(3)	(4)
	<45 w. days	<45 w. days	ln(w. days)	ln(w. days)
Scorecard	0.0449 (0.0699)	0.0604 (0.0509)	-0.141 (0.168)	-0.0250 (0.0900)
Start month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	1,367	1,367	1,367	1,367
Clusters	108	108	108	108
Control mean	0.44	0.51	102.09	69.33

This table compares the effects estimated using the administrative and survey data. The regression specification is the same as in Table I.2. All observations are applications matched between the survey and administrative data. Columns (1) and (3) show the results estimated using the matched administrative data. Columns (2) and (4) show the results estimated using the matched survey data. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Appendix Section B.2.4.



Table I.A.16: Effect by Date of E-Governance System Installation

	(1) <45 w. days	(2) IHS(w. days)	(3) ICW index
Scorecard	0.0568** (0.0275)	-0.117** (0.0596)	0.122** (0.0593)
Treat x installation date	-0.00303 (0.00409)	0.00197 (0.00895)	-0.00467 (0.00887)
E-governance installation date	-0.0109 (0.00812)	0.0253 (0.0179)	-0.0246 (0.0177)
Start month FE	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes
Observations	1,050,924	1,050,924	1,050,924
Clusters	311	311	311
Control mean	0.56	65.64	
Fraction imputed		0.06	
Fraction zero		0.003	

This table shows differences in the effects of the scorecards between offices that had the e-governance system installed during different time periods. The e-governance installation date is the date the first application was made using the e-governance system made in that office. The unit of the installation date variable is months and the variable is measured relative to the weighted mean of installation dates among the offices in the sample. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Appendix Section C.1.1. Discussed in Appendix Section C.1.

Table I.A.17: Robustness to Excluding Applications Potentially Affected by Applicant Survey

	(1) <45 w. days	(2) IHS(w. days)	(3) ICW index
Scorecard	0.0683** (0.0293)	-0.168** (0.0668)	0.140** (0.0593)
Start month FE	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes
Observations	545,742	545,742	545,742
Clusters	310	310	310
Control mean	0.52	79.75	-0.00
Fraction imputed		0.05	
Fraction zero		0.003	

This table shows the results from Table I.2 when restricting the sample to applications that were made either in offices where the applicant survey did not take place, or made one month or more before the start of the applicant survey. Hence these results are highly unlikely to have been affected by the survey activities. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Appendix Section C.1.

Table I.A.18: Effect on Applications Received and Land Size

VARIABLES	(1) ln(Apps. rec.)	(2) ln(Apps. rec.)	(3) ln(Land size)	(4) ln(Land size)
Scorecard	-0.0517 (0.0737)		0.0229 (0.0674)	
Scorecard x Overperform baseline		-0.0577 (0.107)		0.0243 (0.0946)
Scorecard x Underperform baseline		-0.0431 (0.107)		0.0214 (0.0979)
Underperform baseline		-0.157 (0.124)		-0.000536 (0.112)
Observations	311	311	1,042,987	1,042,987
Stratum FE	Yes	Yes	Yes	Yes
Start month FE			Yes	Yes
Weighted by office			Yes	Yes
Clusters			311	311

This table shows the effect of the scorecards on the number of applications received and the land size of those applications. In Columns (1) and (2) observations are at the office level. In Columns (3) and (4) the observations are at the application level. Data contains all applications made between one month before the start of the experiment started and 45 working days before the experiment ended (13 Aug 2018 - 20 Jan 2020). Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Appendix Section C.3.

Table I.A.19: Effect on Rejections

	(1)	(2)
	Rejection	Previously rejected
Scorecard treatment	-0.00810 (0.0196)	0.0215 (0.0203)
Start month FE	Yes	Yes
Stratum FE	Yes	Yes
Weighted by office	Yes	Yes
Observations	1,050,924	3,215
Clusters	311	112
Control mean	0.29	0.06

This table shows the effect of the scorecards on rejections of applications for land record changes. Column (1) shows the effect of the scorecards on the fraction of applications rejected in the administrative data. Column (2) shows the effect on the fraction of applicants surveyed who was returning after having had their application rejected, which is a proxy for an incorrect rejection. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. \*\*\* $p < 0.01$ ; \*\* $p < 0.05$ ; \* $p < 0.1$ . Discussed in Appendix Section C.3.2.

Table I.A.20: Effect on Applicant Satisfaction

	(1)	(2)
	Satisfaction	Satisfaction
Scorecard	-0.0477 (0.0612)	
Scorecard x Overperform		-0.117 (0.0827)
Scorecard x Underperform		0.00746 (0.0897)
Overperform baseline		0.200** (0.0975)
P-value sub-group diff.		0.32
Start month FE	Yes	Yes
Stratum FE	Yes	Yes
Weighted by office	Yes	Yes
Observations	3,018	3,018
Clusters	112	112

This table shows the effect of the scorecards on applicants stated satisfaction transformed from a five-point scale into standard deviations away from the control group mean. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1. Discussed in Appendix Section C.5.

## Part II

# The Effects of Social Movements: Evidence from #MeToo (Joint with Ro'ee Levy)

## 1 Introduction

Societal changes are often associated with movements advocating for new norms and behaviors. For example, the increase in women's labor force participation, the shift in attitudes toward LGBTQ individuals, and the increased concern for the environment all happened in conjunction with social movements advocating for these changes. Despite the importance of these changes, it is difficult to establish if these movements are the drivers of change or if they are caused by external factors that would have led to societal changes regardless of the movements. In this paper, we focus on the MeToo movement and estimate its effect on reporting sexual crime to the police.

The MeToo movement started on October 15, 2017 and was exceptionally effective in rapidly increasing awareness around sexual misconduct. We show that the movement dramatically increased search interest in and news coverage of sexual misconduct. While the movement spread internationally, there was large variation in its strength across countries. We exploit the variation in the strength of the movement, along with the fact that it started almost instantly, to identify its causal effect.

We focus on reporting sexual crimes because underreporting of sexual crimes is a major social problem directly related to the goal of the MeToo movement—sharing one's story and breaking the stigma surrounding being a victim of sexual misconduct. In addition, reporting a sexual crime is a high-stakes decision as it can come with substantial costs in terms of the victim's time, social stigma, the negative experience of reliving the trauma, and the risk of reprisals. Hence, using the number of crimes reported to the police as the outcome variable is a high bar for the types of behaviors that the MeToo movement might have changed.

We construct a dataset on the number of crimes reported to the police by quarter in 30 OECD countries, covering 88% of the OECD population. We identify the effect of the MeToo movement

using a triple difference strategy comparing countries with weak and strong MeToo movements, sexual and non-sexual crimes, and the pre and post periods. We classify countries as having a strong or weak MeToo movement based on Google search interest for terms related to the MeToo movement. We find that the MeToo movement increased the number of reported sexual crimes by 10% during the first six months of the movement.<sup>39</sup> While countries with strong MeToo movements are different from countries with weak movements, we show that the two sets of countries have similar pre-trends for the difference between sexual crime and non-sexual crime.

We confirm the reliability of the result by performing placebo tests where we estimate the effects of fictional MeToo movements set in each of the six-month periods from the second quarter of 2010 to the third quarter of 2017 (Q2 2010-Q3 2017). The effect we find is larger than all the placebo estimates. We also show that the result is robust to various specifications and alternative measures of the strength of the MeToo movement. While the point estimates are similar, our power is limited and the standard errors increase in some of the robustness tests. We also find an effect when employing the matrix completion method (Athey et al., 2017), which uses more flexible patterns in the data to create a counterfactual for the number of sexual crimes reported had there been no MeToo movement. Furthermore, the result is similar when instrumenting the strength of the MeToo movement with the share of the population speaking English.

To measure the persistence of the effects, we focus on the countries with an initially strong MeToo movement and use a difference-in-difference strategy comparing sexual crime with all other crimes over time.<sup>40</sup> We find that the movement had a persistent effect, and estimate a strong effect on reporting at least 15 months after it started.

The `international_dataset` allows for the strongest identification strategy, but it lacks details on the crimes reported. To better understand the mechanisms underlying the effect of the MeToo movement, we use detailed incident-level data from the US at both the national and city level. The US national dataset is collected from the FBI National Incident-Based Reporting System (NIBRS) and covers approximately 30% of the US population. The city dataset, which includes additional covariates and covers more crime categories, is collected from seven large US cities. The US lacks

---

<sup>39</sup>The estimate of the effect is 10 log points, which equals a 11% increase. For simplicity, we describe the effects in log points as percentage changes throughout the paper, although this slightly understates the magnitude of the results.

<sup>40</sup>We use this strategy instead of our main triple difference specification since the MeToo movement became more prominent over time in several of the countries where it was initially weak.

substantial geographic heterogeneity in the strength of the MeToo movement; therefore we employ a difference-in-difference strategy comparing sexual crimes to all other crimes over time. We find that the MeToo movement increased the number of sexual assaults reported in the US by 8% in the first six months after the movement started, and the effect is stronger in the city dataset which also includes sexual harassment.

We present three additional findings based on the US data. First, the movement had a larger effect on crimes that were reported at least a month after they occurred. However, the effect on crimes that were immediately reported is also strong and statistically significant, implying that even if part of the effect of the movement is due to the reporting of a stock of old crimes, the movement also increased reporting of the flow of new crimes. Second, we do not find evidence for the claim, commonly made in media reports, that the MeToo movement mainly affected white women of high socioeconomic status. However, we do find that the movement has a stronger effect among female victims and in counties with a smaller share of Trump voters. Third, we show that in the long run, the movement also led to an increase in the number of arrests made for sexual crimes, suggesting that reporting led to positive externalities.

We discuss several possible mechanisms explaining the increase in reporting. A potential interpretation of the results is that the MeToo movement increased the *incidence* of sexual crimes. By focusing on crimes that were committed before the start of the MeToo movement, while still including crimes reported after the start of the movement, we show that an increase in the incidence of sexual crimes cannot explain the effect we find, and therefore, we conclude that the movement increased the propensity to report crime. The MeToo movement did not lead to major immediate changes in laws or government institutions and therefore legal changes could not be driving the increase in reporting. The mechanism that we have the strongest evidence for is a change in social norms and information. Awareness of the problem of sexual misconduct increased after the MeToo movement started, suggesting that awareness may have led to additional reporting.

The results are related to three different streams of literature. First, we contribute to a long debate among social scientists on whether social movements have any political influence (Burstein and Sausner, 2005). In a review of the topic, Amenta et al. (2010) state that “[t]he disagreement on this basic issue is wide. Some ... hold that social movements are generally effective and account for most important political change. Others ... argue that social movements are rarely influential.”



Papers in this field often document a correlation between a movement's activity and an outcome, such as congressional attention (e.g., Baumgartner and Mahoney, 2005), but do not necessarily identify causal effects. A smaller literature focuses on the causal effects of political protest, a specific tactic often employed by social movements. This literature has shown that protests can mobilize people and change voting behaviors, but that violent protest may also cause a political backlash leading to less political support and subsequent electoral defeat (Madestam et al., 2013; Wasow, 2020). We bridge these literatures by identifying the causal effect of an important social movement. An additional contribution of our paper is that we do not focus on political outcomes, which are typically studied, but rather show how a social movement can affect costly *personal* decisions. Studying the effects of social movements on such decisions is important since many social movements focus on changing norms or individual behavior and not only official policy. Personal decisions also often carry high stakes for the individual, which may make them more difficult to change than voting decisions, for example.

A second contribution to the literature is showing how norms can rapidly change. It is well established that social norms, and especially gender norms, have strong effects on behavior (e.g., Alesina et al., 2013; Bertrand et al., 2015; Charles et al., 2018). However, there is still limited understanding of how social norms change. Several studies have shown that popular culture and education can affect norms and behavior (e.g., Jensen and Oster, 2009; Chong and Ferrara, 2009; La Ferrara et al., 2012; Dhar et al., 2018; Banerjee et al., 2019; Green et al., 2020). There are also well-documented examples of how deceptive practices can lower trust toward certain institutions and change behavior (Alsan and Wanamaker, 2017; Martinez-Bravo and Stegmann, 2018). A recent literature based on theory, as well as information interventions, argues that social norms can "unravel" when individuals start expressing their personal beliefs (Bursztyn et al., 2017, 2018; Sunstein, 2019). We contribute to this literature by demonstrating in an important real-world setting that norms can shift quickly and change important behaviors as awareness to a social issue rises.

This paper also contributes to the literature on reporting gender-based violence by showing that awareness-raising campaigns can have a large effect on the reporting of sexual crimes. Previous studies have shown that the election of female politicians and the integration of women into the police force increased the reporting of crimes toward women (Iyer et al., 2012; Miller and Segal,

2019), and that a high-profile rape and murder case increased reporting of sexual crimes in India (Bhatnagar et al., 2019; McDougal et al., 2018). Public campaigns increasing awareness is a common strategy to increase reporting.<sup>41</sup> However, there is limited evidence on the effects of such campaigns. The MeToo movement can be seen as a particularly successful attempt to raise awareness. To the best of our knowledge, this is the first rigorous evidence on the effects of the MeToo movement on reported sexual crimes and thus demonstrates that increasing awareness can be effective in increasing the reporting of sexual crimes, even in the absence of changes to laws and government policies.<sup>42</sup>

The rest of the paper is organized as follows. Section 2 discusses the underreporting of sexual crime and describes the MeToo movement in more detail. Section 3 describes the international\_data, our identification strategy, and provides evidence for the effect of the movement. Section 4 describes the US data and provides results on heterogeneity as well as the effect on arrests. Section 5 provides evidence on which mechanisms the effect operated through and interprets the overall results, and Section 6 concludes.

## **2 Underreporting of Sexual Misconduct and the MeToo Movement**

### **2.1 Reporting of Sexual Misconduct**

Underreporting of sexual misconduct is a serious problem. Figure II.1 shows that among eight countries that reported data to the UN Sustainable Development Goals (Australia, Canada, France, Iceland, Korea, Mexico, Sweden, and the US), on average 15% of sexual assaults were reported to the police between 2010 and 2017, compared to 35% of other assaults.<sup>43</sup> Underreporting decreases social welfare by reducing the probability that perpetrators are held accountable. Thus it may increase the incidence of sexual misconduct because repeat offenders are not prevented from committing additional crimes, and future offenders are not deterred. Indeed, Green et al. (2020)

---

<sup>41</sup>For example, the largest US-based anti-sexual violence organization RAINN spends 27% of its budget on educating the public.

<sup>42</sup>Rotenberg and Cotter (2018) present descriptive statistics showing that sexual crimes reported increased in Canada after the MeToo movement started. Recent research has also examined other effects of the movement. Lins et al. (2020) show that firms with women among the five highest-paid executives earned excess returns of 1.6% in the second half of October, 2017, after the MeToo movement started. Luo and Zhang (2020) show that producers associated with Harvey Weinstein were more likely to work with female writers after the movement.

<sup>43</sup>In the US, 34% of sexual crime victims stated that the crime is known to the police, compared to 47% of victims of other violent crimes. These figures are based on 2010-2017 National Crime Victimization Survey microdata.

and Iyer et al. (2012) provide suggestive evidence showing that increases in reporting reduce the incidence of gender-based violence.

Reporting a sexual crime to the police is a high-stakes decision for the victim. The process of reporting and attending hearings has monetary costs such as lost income, childcare, and travel costs (Morabito et al., 2019). Moreover, reporting a sexual crime forces the victim to repetitively relive the experience by giving detailed accounts of the crime. Reporting is especially hurtful for victims whose account of the event is not believed by law enforcement officials (Spohn and Tellis, 2012). Furthermore, reporting a crime may lead to reprisals by the offender or the community shared by the victim and the offender. According to National Crime Victimization Survey Data, 17% of sexual crime victims who did not report the crime to the police cite fear of reprisals as a reason for not reporting the crime, while the same figure for victims of other violent crimes is 7%.

In this paper, we focus on reporting sexual crimes to the police. However, the MeToo movement also highlighted cases of sexual misconduct that do not constitute a criminal offense but still have negative welfare consequences, such as cases of workplace sexual harassment (Hersch, 2011; Folke et al., 2020). Furthermore, a victim has a range of possible actions to take in response to sexual misconduct. Reporting to the police is probably one of the actions with the greatest consequences. It is therefore likely that if reports to the police increased, other lower-stakes behaviors changed as well. Indeed, there have been anecdotal reports of an increase in the number of calls to helpline centers following the MeToo movement.<sup>44</sup> Therefore, the effects we find on reporting crime to the police are probably a subset of the overall behavioral effects of the movement.

## **2.2 The MeToo Movement**

The MeToo movement went viral on October 15, 2017, after the Harvey Weinstein sexual misconduct allegations, when a tweet by Alyssa Milano encouraged people who had been sexually harassed or assaulted to write "Me too" on social media.<sup>45</sup> The movement uncovered a large number of sexual misconduct cases, and within a year, more than 200 high-profile men had been ousted from their

---

<sup>44</sup>Chiwaya, Nigel - New data on #MeToo's first year shows 'undeniable' impact. NBC News. Oct 11, 2018. Online: <https://www.nbcnews.com/news/us-news/new-data-metoo-s-first-year-shows-undeniable-impact-n918821>

<sup>45</sup>The phrase "Me Too" was first used by Tarana Burke in 2006, but widespread usage only started after October 15, 2017.

positions in the US alone.<sup>46</sup>

The MeToo movement provides a setting particularly well suited to the study of the effects of social movements on behavior for four reasons. First, the movement was very effective in drawing attention to sexual harassment and sexual misconduct. While the movement started in the US, its effect quickly spread to other countries. Figure II.2 shows that in the OECD, mean Google search interest for MeToo and for sexual misconduct (sexual harassment and sexual assault) increased substantially immediately after the start of the MeToo movement. In the year following the start of the movement, there was an unprecedented increase of 85% in search interest in sexual misconduct compared to January 2010-September 2017. The salience of the movement was widespread across mediums. In the US, approximately eight months after the movement started, 65% of social media users stated that some or a great deal of the content they see on social media is about sexual harassment or assault.<sup>47</sup> Consumers of mainstream media were also likely to encounter the movement. Appendix Figure II.A.1 shows that among four major US newspapers, coverage related to sexual assault and sexual harassment increased substantially after the movement started and remained much higher than the average coverage before the movement started for at least nine months.

Second, there was large variation in the strength of the movement between countries, as shown in Figure II.3. The OECD country in the 75th percentile in terms of MeToo search interest had a 651% larger interest in the MeToo movement in October 2017, compared to the country in the 25th percentile. This allows us to identify the causal effect of the MeToo movement by comparing changes across countries. Third, one of the main objectives of the MeToo movement, increasing reporting of sexual misconduct, is an outcome for which there is high-quality administrative data across many countries. Fourth, while the MeToo movement had a big impact on the public discourse, it did not result in immediate widespread changes to laws or government institutions. This allows us to attribute the short-run effect we find to changes in information and social norms, where norms are broadly defined to include the norms of victims, firms, and government employees, such as police officers, but exclude any changes to laws or government policy.

While the MeToo movement was very successful in raising awareness, it is by no means unique.

---

<sup>46</sup>The New York Times - #MeToo Brought Down 201 Powerful Men. Nearly Half of Their Replacements Are Women. October 23, 2018. Available online: <https://www.nytimes.com/interactive/2018/10/23/us/metoo-replacements.html>

<sup>47</sup>Pew Research Center American Trends Panel Wave 35.

In recent years, several social movements such as Black Lives Matter and March for Our Lives have had similar success in raising awareness about their causes (Pew Research Center, 2018). Social media has enabled new social movements to raise awareness at a larger scale, within shorter time spans, and with almost no organizing structure.<sup>48</sup> However, little is known about the effects of these modern social movements that are often disconnected from party politics and do not use traditional organizing techniques such as strikes or publishing lists of demands.

### **3 Identifying the Effect of the Movement: Analysis of International Data**

#### **3.1 Data**

##### **3.1.1 Outcome: Reported Crimes**

We build a dataset with quarterly data on the number of crimes reported in 30 OECD countries representing 88% of the OECD population.<sup>49</sup> We include in our sample countries that have quarterly, or more frequent, data available, disaggregated by sexual crimes and non-sexual crimes. For 24 of the countries, the time period that a crime is counted in is based on the date the crime was reported to the police, for the remaining countries it is based on when the crime occurred or some combination of the two. We separately obtain data available from the start of 2010 until the end of 2018 for each country. We harmonize the data by manually classifying offense categories as sexual crimes or non-sexual crimes for each country. We define sexual crimes as all forms of sexual assault and sexual harassment and define non-sexual crimes as all other crimes. When possible, we exclude crimes of sexual nature that were not the focus of the MeToo movement, such as incest, human trafficking, and pornography. For more details on crime classification and OECD data collection, see Appendixes A.1 and A.2, respectively.

---

<sup>48</sup>Enikolopov et al. (2019) show how social media facilitated protests in Russia. Acemoglu et al. (2017) show that street protests, but not Twitter protests, can reduce the valuation of politically connected firms and may serve as a check on political rent-seeking. There is also a literature on how different technologies enable the diffusion of social movements (e.g., Christensen and Garfias, 2018; García-Jimeno et al., 2018).

<sup>49</sup>See Appendix Table II.A.5 for a list of the countries and data sources.

### 3.1.2 Strength of the MeToo Movement

We use monthly Google Trends data on search behavior from 2010-2018 to create a proxy for the strength of the MeToo movement in each OECD country.<sup>50</sup> The primary measure is based on the proportion of total Google searches for the "topic" of the MeToo movement. Google defines a search for a topic as any search query including a phrase directly linked to the topic in any language. While Google search interest is an imperfect measure of the MeToo movement's strength, it provides a consistent measure of the movement's strength among a majority of the population, as Google is the dominant search engine in all of the countries in our data.<sup>51</sup> Appendix A.3 provides more details on how the Google Trends data was processed.

We define *immediate interest* as the interest in the MeToo movement during October 2017, the month the MeToo movement started. In our main specification, we categorize a country as having a *strong* MeToo movement if the immediate interest is above the OECD median and a *weak* MeToo movement if the immediate interest is below the OECD median. Figure II.3 shows the immediate interest of each OECD country, highlighting the countries for which we have crime data and indicating which of these countries we classify as having strong and weak MeToo movements. Appendix Figure II.A.2 confirms the validity of our primary measure for the strength of the MeToo movement by comparing it with survey data on the fraction of the population that has heard of the MeToo movement (YouGov, 2019). Even though the survey took place in February-March 2019 and our measure is based on data from October 2017, there is a strong correlation of 0.69 between the two measures.

## 3.2 Empirical Strategy

Our main empirical strategy to measure the causal effect of the MeToo movement on sexual crime reported to the police is a triple-difference strategy where we use the difference over time, across

---

<sup>50</sup>Caputi et al. (2019) show that the MeToo movement affected Google search interest in the US.

<sup>51</sup>In October 2017 among the countries in our sample, the mean of Google's market share of searches was 90%, while the minimum was 66%. Authors' own calculations using data from [gs.statcounter.com](https://gs.statcounter.com).

countries, and between sexual crimes and non-sexual crimes:

$$y_{itc} = \beta_1 Post_t \times StrongMeToo_c \times SexCrime_i + \beta_2 Post_t \times SexCrime_i + \beta_3 Post_t \times StrongMeToo_c + \beta_4 Post_t + \beta_{5,ic} Trend_t + \gamma_{i,c,q(t)} + \varepsilon_{itc} \quad (8)$$

- $y_{itc}$  is the natural logarithm of the number of reported crimes of type  $i$ , in quarter  $t$ , in country  $c$
- $Post_t$  is an indicator for Q4 2017 (when the MeToo movement started) and later quarters
- $StrongMeToo_c$  is an indicator for whether country  $c$  had a strong MeToo movement
- $\beta_{5,ic} Trend_t$  controls for differential linear time trends by the full interaction of country and crime category
- $\gamma_{i,c,q(t)}$  controls for the full interaction of country, calendar quarter, and crime category fixed effects

The regression is unweighted and uses standard errors that are clustered at the country-by-crime category level because that is where the MeToo movement varies.<sup>52</sup> Our identifying assumption is that without the MeToo movement, the difference between sexual crimes and non-sexual crimes would have changed in the same way from the pre-period to the post-period (after controlling for crime and country-specific seasonality and for differential linear time trends) in the countries with strong and weak MeToo movements. For an omitted variable to explain the results, it would have to have a non-linear change after October 2017 that affects the number of reported sexual crimes more than it affects reported non-sexual crimes among countries where the MeToo movement was strong, as compared to countries where it was weak. While the strength of the MeToo movement is not random, we have no reason to believe it is correlated with an omitted variable affecting sexual crimes differentially in the post period.

---

<sup>52</sup>The standard errors clustered at the country level are smaller so choosing to cluster the standard errors at the country-by-crime category level is more conservative.

### 3.2.1 Time Frame of Analysis

In Section 3.3, we focus on the effects of the MeToo movement in the short run, defined as the first six months of the MeToo movement. In Section 3.4, we test if the effect is persistent over time. There are two main reasons for separating out the short-run effects. First, the first six months is the time period when there exists a substantial difference in interest between countries with a strong movement and countries with a weak movement. Therefore, this is the only period in which we can employ our triple-difference empirical strategy. Appendix Figure II.A.3 shows the convergence of interest over time between the countries that we classify as having strong and weak MeToo movements. Second, during the initial six-month period there were, to the best of our knowledge, no changes to laws governing sexual crimes in any of the countries in our sample. After the initial six-month period, some laws concerning sexual crime changed in at least three countries, probably as a result of the MeToo movement. Therefore, in the first six months, we can interpret the effect as being driven by a change in social norms or information.

### 3.3 Results

Table II.1 shows that the MeToo movement increased the reporting of sexual crimes. Column (1) uses data only on sexual crimes to show a difference-in-difference estimator over time and between countries with strong and weak MeToo movements. Column (2) uses all 30 countries and shows a difference-in-difference over time and between sexual and non-sexual crime. While the two columns use different sources of variation, they both find statistically significant effects of 11% and 7%, respectively. It is not surprising that Column (2) finds a smaller effect than Column (1) since it estimates the average effect for countries with both strong and weak MeToo movements. Column (3) estimates the effect from Column (2) separately for countries with strong and weak movements and shows that the effect is driven by the countries that had a strong MeToo movement. These countries had an effect of 12%, while the effect was only 2% among countries with weak MeToo movements. Finally, Column (4) shows the results from our main triple-difference specification described in Equation 8. Here the coefficient of interest is that on  $Post_t \times StrongMeToo_c \times SexCrime_i$  and we find an effect of 10%, statistically significant at the 10% level. Note, that in this column, countries with weak MeToo movements serve as a control group. If the MeToo movement had some effect in



these countries, the estimate is a lower bound for the total effect of the movement.

In Columns (3) and (4) of Table II.1, the coefficient on  $Post_t \times StrongMeToo_c$  can be interpreted as a difference-in-difference estimate of the effect of the MeToo movement on non-sexual crimes, using variation between countries and over time. Since we do not expect the MeToo movement to affect the number of non-sexual crimes reported, this coefficient can be used as a placebo test. We estimate the coefficient to be close to zero, which confirms that our estimate of the MeToo movement's effect is not influenced by differential trends in non-sexual crime reporting between countries with weak and strong movements.

To illustrate the triple-difference estimator, we present the raw data visually in Figure II.4. Sub-figure II.4a shows the number of sexual crimes reported, indexed to be 100 in Q3 2017, and averaged across the countries with strong and weak MeToo movements.<sup>53</sup> A clear seasonality is observed in the time lines, where the fourth quarter of each year tends to see a decrease in the number of sexual crimes reported. This is true for both strong and weak MeToo movement countries until Q4 2017, when the number of reported sexual crimes stays flat in the countries with a the strong MeToo movement , while the countries with a weak movement experience the typical decline. In Q1 2018, the number of reported sexual crimes in countries with strong and weak MeToo movements continues to diverge. Sub-figure II.4b shows that this differential increase in reported crimes for the countries with strong MeToo movements did not happen for non-sexual crimes. The figures also shows that the strong and weak MeToo movement countries may have somewhat different pre-trends for sexual and non-sexual crimes. In our main specification, we control for linear time trends, and hence, these differential trends do not drive the effects as measured in Table II.1. Furthermore, Sub-figure II.4c shows that there are no differential pre-trends in the difference between the sexual and non-sexual crime indexes displayed in Sub-figures II.4a and II.4b, , while there is a substantial divergence between countries with strong and weak MeToo movements after the start of the MeToo movement.

Using a continuous estimate of the strength of the MeToo movement, we estimate the elasticity of crimes reported to the national interest in the MeToo movement. Replacing the  $StrongMeToo_c$  term in Equation 8 with the inverse hyperbolic sine transformation (IHS) of the immediate search

---

<sup>53</sup>To make the average numbers more comparable over time, we shorten the data series to start in 2012, since we lack data for many countries before 2012.

interest in the MeToo movement provides an estimate of the effect of an IHS point increase on the log of reported sexual crimes.<sup>54</sup> This regression yields an estimate of 0.05. However, this estimate is not statistically significant and it should be interpreted with caution because Google searches for the MeToo topic is a noisy proxy for the underlying interest in the MeToo movement and thus the estimate probably suffers from attenuation bias.<sup>55</sup>

### 3.4 Persistence of the Effect over Time

Was the effect of the MeToo movement driven by a short-term increase in the salience of exposing sexual crime or did the movement change the underlying social norms leading to a lasting effect on behavior? To estimate the long-term effects, we cannot use the triple-difference estimator, because in some of the countries where a MeToo movement was initially weak, it gained traction and became stronger after October 2017, as shown in Appendix Figure II.A.3. This means that when measuring long-term effects, our counterfactual will become contaminated in later periods. We use two alternative strategies instead. First, we use the difference-in-difference specification over time and by crime type and focus only on countries where we know that the movement started in October 2017. Second, in Appendix Section B.1, we exploit the gradual spread of the movement and allow the MeToo movement to start at different time periods in different countries to estimate the effect over time. Using both methods we find that the movement's effect was persistent.

Table II.2 uses data from the countries with a strong MeToo movement to measure the persistence of the effect over time. Column (1) shows that the average effect for the first five quarters after the movement started is estimated to be 10%. Column (2) shows that the effect is relatively stable until the end of our data, 15 months after the movement started. The effect stays between 8% and 12%, and there is not a pattern of a continuous change in the effect over time.

---

<sup>54</sup>We use the IHS transformation instead of the natural logarithm since for one country the estimated interest is negative, but very close to zero.

<sup>55</sup>When instrumenting the interest in the MeToo movement using the fraction of the population that speaks English, the point estimate increases substantially, suggesting attenuation bias is affecting the estimate. For a more detailed description of the instrumental variable approach see Section 3.6.

### 3.5 Placebo Tests

We conduct a set of placebo tests to further assure that the MeToo movement is driving our result and not some other mechanism, such as non-linear differential trends between countries with strong MeToo movements compared to those with weak movements. Figure II.5 presents placebo tests setting the start of the MeToo movement in every second quarter from Q2 2010 to Q4 2017 and then measuring the effect over six-month periods.<sup>56</sup> We estimate the effect of these placebo MeToo movements using the triple-difference specification from Equation 8, just as we do in our main specification in Column (4) of Table II.1. Of the 15 placebo tests, only one is statistically significant at the 10% level. The actual effect of the MeToo movement (Q4 of 2017) has a larger absolute coefficient than any of the 15 placebo tests.

### 3.6 Robustness Checks

Table II.3 shows that our primary triple-difference estimator is robust to using different time periods, alternative regression specifications, alternative empirical strategies, and to most alternative definitions for the strength of the movement. Row (1) repeats the main estimate from Column (4) of Table II.1. Row (2) shows the effect of the MeToo movement during its first quarter by restricting the sample to end in Q4 2017. Note that the first two weeks of this quarter could not have been affected by the movement and therefore this result probably underestimates the effect of the movement. Row (3) shows the effect of the MeToo movement during the first three quarters of the movement. All the effects range from 6% to 10%.

Rows (4)-(6) estimate the effect with different measures of the strength of the MeToo movement. Row (4) shows that the result is robust to using Google searches for the MeToo topic between October 2017 and March 2018, the same period for which we measure the number of reported crimes. Row (5) uses the sum of the Google search interest in the topics of sexual assault and sexual harassment.<sup>57</sup> Using this noisier measure of MeToo strength produces a smaller estimate. Row (6)

---

<sup>56</sup>We set the start date in every second period to avoid having two adjacent estimates using overlapping data and thereby introducing a mechanical autocorrelation. When estimating the placebo effect for every quarter, we still find that only one placebo test has a statistically significant effect at the 10% level and that the actual effect of the MeToo movement has a larger absolute coefficient than any of the 31 placebo tests.

<sup>57</sup>Since there was a meaningful interest in these topics even before the start of the MeToo movement, we use the increase in search interest at the start of the MeToo movement after controlling for linear trends and monthly fixed effects in each country separately. This allows us to parse out pre-MeToo levels of interest, linear trends, or seasonality, which are not indicative of the strength of the MeToo movement.

uses a survey measure of the fraction of the population that has heard of the MeToo movement in February-March 2019 (YouGov, 2019). The analysis is conducted for the 12 countries in our sample where the survey was conducted. This analysis yields a point estimate similar to our main estimate.

Row (7) shows the result of a regression weighted by the population of each country.<sup>58</sup> Using these weights changes the interpretation of the estimate from the average effect of the MeToo movement on the number of sexual crimes reported in countries that had a strong MeToo movement to the average effect of the MeToo movement on the population in the countries that had strong MeToo movements. This effect is estimated to be 12% and is more precisely estimated than our main estimate since we put more weight on countries with a large population that on average have a more stable quarter-to-quarter number of crimes reported. While most of our data is based on the date crimes were reported to the police, some of the data is based on the date crimes occurred. This may bias the results as crimes that occurred before the start of the MeToo movement could also be affected by the movement. Row (8) shows the results of our main specification including only data based on the date a crime was reported and confirms that differences in this reporting practice do not drive the results.

To ensure that our specification of the outcome variable is not driving the result, Row (9) shows the result when using the number of crimes reported as an outcome variable, whereas our main specification uses the log of crimes reported. We normalize the number of crimes reported to average one in the year before the start of the MeToo movement in each country by crime type category. The estimated effect is an 11% increase over the baseline year (Q4 2016 - Q3 2017). Row (10) shows the result is robust to using a negative binomial regression with the count data of crimes reported as the outcome variable.

Row (11) analyzes the data using the matrix completion method which creates a counterfactual for the number of sexual crimes that would have occurred in countries that had a strong MeToo movement, based on flexible patterns in the data.<sup>59</sup> Despite using a very different empirical strategy the estimated effect is qualitatively similar to our main estimate. A potential problem with our main specification is that reverse causality could bias the results if an increase in sexual crime

---

<sup>58</sup>We use UN population data for 2015 from the 2017 revision of World Population Prospects.

<sup>59</sup>Each row in the matrix is a crime\*country category and each column in a year-quarter (for example a control category may be damage to property in Ireland and a treated category may be sexual harassment in Iceland). For more details on the method see Section 4.4).

reporting increased the interest in the MeToo movement and thus affected the classification of strong and weak movements. To rule out such a mechanism we instrument having a strong MeToo movement with the fraction of the population speaking English.<sup>60</sup> Since an increase in reported sexual crimes could not have affected the fraction of the population speaking English, this estimate should not suffer from reverse causality bias. Row (12) shows that our main result is robust to using this two-stage least squares regression.<sup>61</sup>

## 4 Heterogeneity and Effect on Arrests: Analysis of US data

To study heterogeneity and mechanisms in the effects of the MeToo movement, we focus on the US since that is where the movement started and because rich incident-level data is available for the US.

### 4.1 Data

We use US data from two sources: the FBI National Incident-Based Reporting System (NIBRS) and more detailed crime data for seven large US cities.

#### 4.1.1 National Data: FBI NIBRS

Law enforcement agencies voluntarily report data on offenses as part of the FBI's Uniform Crime Reporting (UCR) Program. Agencies have been gradually shifting from reporting summary statistics of the most severe offenses to reporting incident-level data using the NIBRS for 52 specific crimes, defined as Group A offenses.<sup>62</sup> By 2017, more than 7,000 agencies covering approximately 30% of the US population reported data using the NIBRS program. In our main specification, we use 2010-2018 NIBRS data aggregated at the state by crime category level for each month. Similarly to the international analysis, we aggregate data into two main categories: sexual crime

---

<sup>60</sup>We use two variables based on Ethnologue data: the share of the population speaking English as a first language and the fraction of the population speaking English. We instrument the interactions of *Post*  $\times$  *Sexual Crime*  $\times$  *Strong MeToo* and *Post*  $\times$  *Strong MeToo* with the the same interactions, where Strong MeToo is replaced with each English speaking measure. See Appendix Section A.4 for a description of the data on English usage.

<sup>61</sup>The Kleibergen-Paap Wald test statistic is 32.

<sup>62</sup>For more details, see the 2019 National Incident-Based Reporting System User Manual. Available online: <https://ucr.fbi.gov/nibrs/nibrs-user-manual>

and non-sexual crime. Group A offenses do not include sexual harassment, therefore our estimates measure the effect only on sexual assaults.

The main advantage of using NIBRS data is that the crime categories and the variables describing each incident are harmonized across law enforcement agencies. This allows us to test for heterogeneous effects by crime type, the characteristics of the victim and offender, and whether an arrest was made. Appendix A.5 provides more details on how the NIBRS data was processed.

#### **4.1.2 Incident-Level Data from Cities**

We collect incident-level data from seven large US cities with a combined population of 16 million: Denver, Kansas City, Los Angeles, Louisville, Nashville, New York City, and Seattle. Our sample consists only of cities that provide incident-level data on all crimes and provide both the date each crime occurred and the date it was reported, along with the crime's approximate location. The seven cities selected are the cities that met our inclusion criteria among the 50 largest US cities.

The city data is used to complement our analysis in three ways. First, information on the location of each incident allows us to analyze heterogeneity in the effect of the MeToo movement by neighborhood. Second, we use the detailed reporting and occurrence dates to analyze heterogeneous effects according to whether the crime was immediately reported. Third, the data includes virtually all crimes reported to the police, and not only the relatively severe offenses covered by NIBRS.<sup>63</sup> Specifically, this allows us to analyze the effect of the movement on sexual harassment, in addition to sexual assault.

We aggregate the incident-level crime data into three main categories: sexual assault, sexual harassment, and non-sexual crime. We manually classify the crime categories for each city separately and exclude crimes that could be indirectly affected by the MeToo movement. In our main specification, we aggregate data at the city by crime category by month level. Appendix A.6 provides more details on how the city data was processed.

## **4.2 Empirical Strategy**

We analyze US data using a difference-in-difference specification over time and by crime type. We do not use a triple-difference strategy, as we do not observe meaningful variation in the strength of

---

<sup>63</sup>There are several exceptions, such as cities excluding crimes related to child abuse cases or unfounded complaints.

the MeToo movement across different regions within the US, as seen in Appendix Figure II.A.4.<sup>64</sup> This is unsurprising as the national media covered the movement and the allegations related to it. Furthermore, the movement generated substantial public discussion in social media, which is not limited to a specific media market. Indeed in a PEW survey from November-December 2017, 92% of Americans reported reading or hearing about recent allegations of sexual harassment and assault.<sup>65</sup>

We use the following regression as our primary specification:

$$y_{itc} = \beta_1 \text{SexCrime}_i \times \text{Post}_t + \beta_2 \text{Post}_t + \beta_{3,ic} \text{Trend}_t + \gamma_{i,c,m(t)} + \varepsilon_{itc} \quad (9)$$

- $y_{itc}$  is the inverse hyperbolic sine transformation of the number of reported crimes of type  $i$ , in month  $t$ , in location (state or city)  $c$ . The inverse hyperbolic sine is used instead of a log transformation because there are months when no crime is recorded for a specific location and crime category
- $\text{Post}_t$  is an indicator for October 2017 and later
- $\beta_{3,ic} \text{Trend}_t$  controls for differential linear trends by the full interaction of location and crime category
- $\gamma_{i,c,m(t)}$  controls for the full interaction of location, calendar month and crime category fixed effects

The specification is similar to our triple-difference specification described in Equation 8 with several differences. First, we aggregate the data at the monthly level, instead of the quarterly level. For each location, we exclude months when no crimes were reported. Second, we use robust standard errors. Since our main specification includes only two crime categories, we cannot cluster the standard errors at the crime category level (where the treatment occurs). Appendix Table II.A.6 uses the same specification, with a finer aggregation of crime categories, which allows us to cluster the standard

<sup>64</sup>While the OECD country in the 75th percentile in terms of search interest had a 651% larger interest in the MeToo movement, compared to the country in the 25th percentile, the same figure for US states was only 47%. Furthermore, the variation between OECD countries was relatively stable over time with a correlation of 0.95 between interest in October 2017 and interest in November 2017, while the same correlation for US states was just 0.34. The low correlation indicates that a large part of the variation in interest between US states is probably due to noise and not actual differences in the strength of the MeToo movement.

<sup>65</sup>Pew Research Center, December 2017 Political Survey.

errors at the crime category level, and shows that the point estimates and standard errors remain similar. In Section 4.4, we show that the results are also robust to an estimation strategy using a finer aggregation of crime categories at the city level and bootstrapping the standard errors. A third difference is that we weight regressions by the average number of crimes that occurred in a location in the pre-period since we are interested in the effect of MeToo on the number of crimes reported and not in the effect of the movement on an average city or state.<sup>66</sup> An additional advantage of weighting the data is that the weights reduce the importance of the aggregation method in our estimates (e.g., whether we aggregate the data by state or county).

#### 4.2.1 Heterogeneity by Demographics

We estimate heterogeneity by the county where the crime occurred using the following regression:

$$y_{itc} = \beta_1 \text{SexCrime}_i \times \text{Post}_t + \beta_2 \text{Post}_t + \beta_3 \text{SexCrime}_i \times \text{Post}_t \times \text{Demog}_c + \beta_4 \text{SexCrime}_i \times \text{Demog}_c + \beta_5 \text{Post}_t \times \text{Demog}_c + \beta_6 \text{Demog}_c + \beta_{7,ic} \text{Trend}_t + \gamma_{i,c,m(t)} + \varepsilon_{itc} \quad (10)$$

The regression is based on Equation 9 with  $c$  now representing a county instead of a city/state and  $\beta_3$  estimating heterogeneous effects by the demographics of the county. Each demographic variable ( $\text{Demog}_c$ ) is constant across time and its weighted mean is subtracted to keep  $\beta_1$ , the estimates for the effect of the MeToo movement, consistent across specifications. Data on county-level income, education, race, and ethnicity is based on the American Community Survey 5-year 2016 estimates. The share of Trump voters in each county is based on the MIT Election Data and Science Lab (2018). We exclude counties with a population of less than 10,000 and county-years where the police agencies reporting data cover less than 85% of the population.

### 4.3 Results

Table II.4 shows that the MeToo movement had a strong and statistically significant effect on crimes reported based on both the NIBRS and the city datasets. Column (1) uses NIBRS data to show that the MeToo movement increased the number of reported sexual assaults in the US by

---

<sup>66</sup>The international analysis regressions in Section 3.5 are not weighted, since in this analysis the treatment occurs at the country level and we are interested in the average effect of the MeToo movement on different sets of countries.



8% in the six months after the movement started. This effect may be smaller than our primary specification based on the international data because the NIBRS dataset includes mostly severe crimes. Column (2) shows that in our sample of large cities, the effects on sexual assault and sexual harassment are approximately 11% and 15%, respectively. As both effects are related to the MeToo movement, in Column (3), we aggregate sexual assault and sexual harassment into one category, labeled sexual crime, which we will focus on throughout the rest of the analysis. We find an effect of approximately 13% on sexual crimes reported in our city sample. To ensure that the effect in one city is not driving the results, we run our main specification separately for each city. Appendix Table II.A.7 shows that the effect is positive for six of the seven cities in our sample and statistically significant for four of the seven cities.

Appendix Table II.A.8 shows that the effect of the movement was persistent in the US and does not decline over time, both based on FBI data and on our sample of cities. One concern with estimating long-run effects is that they can potentially be affected by depletion in the stock of old crimes, and not due to a change in the effect on the propensity to report crime. The city data allows us to mitigate this concern by focusing only on the flow of crimes that were reported within a month after they occurred. Column (5) shows that the results are similar when focusing only on new crimes reported.

#### **4.3.1 Heterogeneous Effects by Report Timing and Crime Type**

Table II.5 tests for heterogeneity by crime type. Column (1) splits the category of sexual crime according to the specific offense type and shows that the MeToo movement had a large effect on the number of rapes reported, the most severe sexual offense category, and on fondling cases. Column (2) shows that the movement had a stronger effect on offenses where the victim was not physically injured. In Column (3) we do not find substantial heterogeneity by whether the victim knew the offender.

Table II.6 uses city data to show that while the MeToo movement had a stronger effect on crimes reported at least a month after they occurred, the movement also affected crimes which were immediately reported. For this analysis, we aggregate crime into three main categories: sexual crimes reported more than a month after they occurred, sexual crimes reported a month or less

after they occurred, and non-sexual crimes, which is the reference category. Column (1) shows that the movement had an effect of 10% on crimes reported within 30 days, and an effect of 22% on crime reported more than 30 days after they occurred. The total effect on all crimes (shown in Column (3) of Table II.4) is similar to the effect on crimes reported within 30 days since only 20% of crimes are reported more than a month after they occur. Column (2) presents the results for the next nine months and shows a declining effect on crimes reported with a lag. This suggests that some of the short-term effect could be due to a stock of old crimes that was exhausted. However, even in the 7-15 months after the movement started, there is a large effect on crimes reported at least a month after they occurred. This long-term effect on crimes reported with a lag could be explained by either a persistent effect on a flow of cases which are not immediately reported or a very large stock of unreported crimes, which is gradually affected by the MeToo movement.

#### **4.3.2 Heterogeneous Effects by Gender, Race, Socioeconomic Status and Political Ideology**

The MeToo movement has been criticized for focusing on white victims of high socioeconomic status and ignoring the experiences of working-class women and women of color (Onwuachi-Willig, 2018). Based on the analysis of victim, offender, county, and neighborhood demographics, we find that the effect was larger for female victims, male offenders, and politically liberal counties. However, we do not find evidence that the MeToo movement mostly affected the reporting of whites or those with high socioeconomic status.

We test for heterogeneous effects among victims by separating sexual assault into sub-categories according to the victim demographics.<sup>67</sup> Column (1) of Table II.7 shows that the movement had a larger effect on female victims than among male victims. This is consistent with the general narrative of the MeToo movement, which tended to focus specifically on female victims of sexual crimes. Column (2) finds a similar effect on black and white victims, and we cannot reject a homogeneous effect across the victim's race.<sup>68</sup> Column (4) repeats the analysis according to the offender's demographics and points to a similar effect among black and white offenders.

Table II.8 shows that the MeToo movement mostly did not have large heterogeneity in the effect

---

<sup>67</sup>For example, when estimating heterogeneous effects by race, the treated categories are sexual assaults of black victims and sexual assaults of white victims, and the reference category is non-sexual crimes.

<sup>68</sup>The NIBRS also includes data on Hispanic ethnicity. We do not find a stronger effect of the movement on individuals who are not Hispanics or Latinos. We do not present the results by ethnicity since the ethnicity could not be identified for 28% of victims and 82% of offenders.

between counties with different demographic profiles, relative to the total effect of the movement. In Column (1) we show the effect of the movement based on our specification in Equation 9, and in Columns (2)-(7) we estimate heterogeneous effects according to each demographic variable as described in Equation 10. Some of the coefficients are statistically significant, but the magnitude of most of the effects is small. While counties with a larger share of college graduates are associated with a slightly larger effect, the difference in the expected effect on reporting between a county in the 75th percentile of the share of individuals with a college education and a county in the 25th percentile is only expected to be 2 percentage points, compared to the average effect of 9%.<sup>69</sup> One exception to the relatively homogeneous effects we find is that the MeToo movement had a smaller effect in counties with a larger share of Trump voters. The difference in expected reporting between a county in the 25th percentile of Trump voters and county in the 75th percentile is 7 percentage points. We emphasize that these estimates are not intended to capture causal effects, but rather to describe which types of counties are associated with a larger increase in reporting sexual crimes during the MeToo movement. Appendix B.2 exploits the more detailed city data to analyze heterogeneity at the neighborhood-level. We do not find substantial heterogeneity by neighborhood demographics.

### 4.3.3 Effect on Arrests

The NIBRS data allows us to test not only whether crime reporting increased, but also whether the movement had an effect on the number of arrests made by the police. The FBI defines an arrest as a case where a suspect is taken into custody based on a warrant or a previously submitted report, arrested on view (without a warrant), or summoned to court.

Table II.9 shows that the movement increased the number of arrests in sexual assault cases, but that this increase is smaller than the effect on reporting. In Column (1), the short-run effect is estimated by aggregating the data into three separate categories: sexual crimes where an arrest was made, sexual crimes where no arrest was made, and non-sexual crimes, which is the control group. In the short run, we find no effects on arrests. One concern with this specification is that the null effect could be explained by a decrease in the arrest rate over time for all crimes. Columns

---

<sup>69</sup>Just like the rest of the US analysis, the average effects and the percentiles are weighted by the mean number of crimes in the pre-period.

(2) and (3) show that the results are robust to using a slightly different specification: we run the regression separately only on crimes where an arrest was made and crimes where no arrest was made so that the control group for each group of sexual crimes is the non-sexual crimes in the same arrest category. Columns (4)-(6) repeat the analysis for the long-run effect over 15 months. In this period, the MeToo movement increased the number of arrests substantially, albeit the increase is still smaller than the effect on the number of crimes cleared.<sup>70</sup>

Why did the MeToo movement not lead to a larger increase in the number of arrests? One possible explanation is that the movement affected mostly the type of cases where the probability of arrest is low. Indeed, as shown in Table II.6, the MeToo movement had a stronger effect on cases reported more than a month after they occurred.<sup>71</sup> However, the movement also had a strong effect on crimes reported within a month, which are far more common. Therefore, the increased reporting of old crimes cannot explain the disproportionately smaller effect on arrests.<sup>72</sup>

In Appendix Table II.A.10, we test whether other observables associated with the crimes affected by the movement could explain the arrest rate. We focus on the police agency where the crime occurred, the type of crime, the age, race, and sex of the victim, whether the victim was injured, the weapon used, the relationship between the victim and offender, and the type of location where the crime occurred. Even though some of these covariates have been shown to be associated with arrests (Lonsway and Archambault, 2012), they do not explain the decrease in the arrest rate. However, there could be unobservables associated with crimes affected by the MeToo movement that are correlated with the likelihood of arresting an offender.

---

<sup>70</sup>In Appendix Table II.A.9 we estimate the effect on cases cleared by the police. A case is cleared if it is associated with an arrest or if the police have sufficient probable cause to arrest a suspect but could not make an arrest for reasons outside their control, including the victim refusing to cooperate, the prosecutor declining prosecution for a reason other than lack of probable cause, the offender being in the custody of another jurisdiction, and the offender being a juvenile. The effects on clearances are similar to the effects on arrests.

<sup>71</sup>Anecdotal evidence suggests that it was difficult to clear MeToo-related cases since they were reported long after they occurred. For example, see Maddaus, Gene - Many Accused, None Prosecuted: Why #MeToo Hasn't Led to a Single Criminal Charge in L.A. *Variety*. September 25, 2019

<sup>72</sup>Based on cities that collect arrest data (Kansas City, LA, and Nashville), the share of cases resulting in an arrest in the pre-period is 10% for sexual assaults reported at least a month after they occurred, compared to 12% for sexual assaults reported within a month. This gap is not large enough to explain the small effect on the number of arrests. We regress whether an arrest was made on the interaction of crimes and the post-period and find that the MeToo movement had a small negative effect on the arrest *rate*, at least in the short run. To test whether this is explained by increased reporting of crimes that were reported at least a month after they occurred, we control for the lag between the occurrence of the crime and its reporting date. The effect stays almost exactly the same when controlling for the lag.

#### 4.4 Robustness: Matrix Completion Method

Our difference-in-difference specification relies on the assumption that other crimes are a suitable control group for sexual crimes after controlling for crime and location-specific seasonality and differential linear time trends. In this section, we relax those assumptions, and instead of estimating an effect based on the standard difference-in-difference specification, we use the matrix completion method and show that the results are robust to the method used.

The matrix completion method (Athey et al., 2017) is used for panel data and is based on a matrix where each row is a unit and each column is a time period. The method attempts to predict the counterfactual outcome for treated units in the post-period. We use the method to create a counterfactual for the expected number of sexual crimes in the post periods, which would have been reported if there was no MeToo movement. The counterfactual matrix is created for all observations, and values are chosen to minimize the sum of squared differences between the actual outcomes and the predicted counterfactual outcomes for observations that were not affected by the movement (non-sexual crimes in all periods and sexual crimes in the pre-periods), with penalization according to the nuclear norm of the predicted matrix. Penalization is required to prevent overfitting, and the regularization parameter is selected through cross-validation. Finally, the average treatment effect is the weighted difference between the actual outcomes and counterfactual outcomes for the treated units in the post-periods. The main advantage of the matrix completion approach is that it is *“able to model more complex patterns in the data, while allowing the data (rather than the analyst) to indicate whether time-series patterns within units, or cross-sectional patterns within a period, or a more complex combination, are more useful for predicting counterfactual outcome”* (Athey, 2018).

We use this method with our city data and define each unit as a crime category by city combination, and each time period as a month. We use the original crime categories defined for each city and do not aggregate crimes to broader categories.<sup>73</sup> We exclude categories for which there was at least one month with no crimes reported. All sexual assault or sexual harassment crimes that occurred on or after October 2017 are considered treated. In total, we have 39 treated groups and 399 control groups. We explicitly control for category and time fixed effects and do not add any

---

<sup>73</sup>For example, indecent exposure in Los Angeles is a row in the matrix and is considered treated for time periods (matrix columns) on or after October 2017. Simple assault in Nashville is an example for a row in the matrix which is untreated in all time periods.

additional controls. We weight each crime group by the number of reported crimes for that crime group in the pre-period.<sup>74</sup>

We find an average treatment effect over six months of 18% and a long-run 15-month effect of 16%. Both effects are significant at the 1% level using standard errors generated by bootstrapping. Appendix Figure II.A.5a shows that the counterfactual created by the method fits the actual outcome well in the pre-period, and Appendix Figure II.A.5b highlights that the treatment effect is relatively persistent.

## 5 Mechanisms and Interpretation

How did the MeToo movement increase the reporting of sexual crimes? In this section, we show that neither an increase in the incidence of sexual crime occurrence nor changes in legislation are likely to be driving the results. We provide evidence that beliefs regarding sexual misconduct changed after the start of the movement. Specifically, increased awareness of the extent of the problem of sexual misconduct may have led to the effect on reporting.

### 5.1 Changes in the Incidence of Sexual Crimes

The effect of the MeToo movement on the number of crimes reported could be driven by an increase in the incidence of crimes (a “backlash effect”) and not an increase in the propensity to report crimes. Ideally, it would be possible to disentangle the effects on incidence and the propensity to report using survey data, such as the National Crime Victimization Survey or the Campus College Survey. However, the MeToo movement could have affected individuals’ definition of sex crimes and decreased the stigma in reporting victimization to a surveyor. Indeed, in two October 2018 surveys by Ipsos, 54% of respondents agreed with the statement “my views on what constitutes sexual harassment have become more clear” (Ipsos, 2017a), and 24% of employed respondents agreed with the statement “The #MeToo movement has made me realize now that I may have been a victim of sexual harassment in the workplace” (Ipsos, 2017b). Furthermore, Dhar et al. (2018) find that a school-based intervention discussing gender equality, including harassment, increased girls

---

<sup>74</sup>The method was estimated using the R package `gsynth` by Yiqing Xu and Licheng Liu. Available online: [https://yiqingxu.org/software/gsynth/gsynth\\_examples.html](https://yiqingxu.org/software/gsynth/gsynth_examples.html)

reporting that they experienced harassment. The authors attribute the effect to increased awareness or destigmatized victimization.

Instead of relying on self-reported survey data, we rule out that an increase in incidence is driving the entire increase in reporting by restricting our analysis to crimes that were committed *before* the start of the MeToo movement and thus their incidence could not have been affected by the movement. Table II.10 uses the data from US cities and includes only crimes that were reported at least three months after they occurred and that were reported by December 2017 (i.e., occurred before the start of the movement in October 2017). The table shows that the MeToo movement had a strong and statistically significant effect on the reporting of crimes that occurred before the movement started. This evidence complements self-reported survey evidence and anecdotal reports of an increase in the propensity to report sexual crimes.<sup>75</sup> Furthermore, we are not aware of any anecdotal reports suggesting that the number of sexual crimes committed increased as a result of the MeToo movement.

## 5.2 Changes to Laws and Government Policy

The MeToo movement could also affect reporting by changing the laws governing sexual crimes, for example, due to an expansion of the behavior classified as illegal. We find evidence against this mechanism, at least in the short term. A report by the International Lawyers Network (2019) shows that among the 11 OECD countries covered by the report, no country made changes to laws governing sexual misconduct between the start of the MeToo movement and the end of Q1 2018, while some introduced legal changes after this date.<sup>76</sup> The lack of legal changes in the immediate aftermath of the MeToo movement is not surprising given that passing legislation is a lengthy process, often taking more than a year.<sup>77</sup>

<sup>75</sup>in October 2018, 62% of Americans stated that if it happened to them, they would be more likely to report sexual harassment now, compare to a year ago (Ipsos, 2017a). Anecdotally, an increase in the propensity to report sexual crimes has been reported in various settings including colleges (Binkley, Collin - MeToo inspires wave of old misconduct reports to colleges. PBS October 13, 2018); the entertainment industry (Maddaus, Gene - Many Accused, None Prosecuted: Why #MeToo Hasn't Led to a Single Criminal Charge in L.A. Variety. September 25, 2019); and with respect to congressional candidates (Godfrey, Elaine, Felton, Lena and Hosking, Taylor - The 25 Candidates for 2018 Sunk by #MeToo Allegations. The Atlantic. July 26, 2018).

<sup>76</sup>To the best of our knowledge only three countries in our data, Iceland, Sweden and the US, changed major laws with respect to sexual crimes between October 2017 and the end of 2018. In Iceland and the US, the earliest changes took effect in Q2 2018 and in Sweden the change took effect in the Q3 2018. Therefore, these changes could not have directly influenced reporting in the first six months of the movement.

<sup>77</sup>An analysis of US laws conducted by USA Today one year after the start of the MeToo movement found that Congress passed no laws related to sexual harassment in the workplace since the movement started. While there was a

### 5.3 Changes in Awareness and Beliefs

Table II.11 uses survey data to show that awareness of sexual misconduct increased after the movement started. We use data from the Views of the Electorate Research Survey since the survey asked a large panel of respondents the same set of questions before and after the start of the movement: in July 2016 and April-May 2018. In contrast to most recent surveys focusing on issues raised by the MeToo movement, the timing of the survey was not affected by the movement. Column (1) shows that agreement with the statement “*sexual harassment against women in the workplace is no longer a problem in the United States*” decreased by 0.14 standard deviations in 2018, compared to 2016.<sup>78</sup> Other surveys also provide evidence for increased awareness (Castle et al., 2020). In a Washington Post-ABC poll in January 2018, 72% of respondents stated that sexual harassment of women in the workplace is a serious problem, compared to 47% in November 2011.

Awareness can affect behavior by decreasing the stigma associated with reporting (Bursztyn et al., 2017), by allowing individuals to coordinate and provide corroborating evidence (Cheng and Hsiaw, 2019), or by aggregating information and encouraging people to report as a form of protest if they learn that sexual assault is a large social problem (Battaglini et al., 2020). Since awareness affects reporting and is affected by it, an initial increase in awareness may lead to a tipping point that further increase reporting and awareness substantially.<sup>79</sup>

We cannot identify the exact mechanism through which awareness affects reporting, but we can explore the mechanisms further using survey data. Column (2) of Table II.11 provides evidence for heterogeneous effects in awareness between men and women. While men’s agreement with the statement that sexual harassment is no longer a major problem decreased by 0.24 standard deviations, women’s agreement decreased by only 0.05 standard deviation and is not statistically different from zero. Interestingly, it seems that a general increase in awareness may have an effect, even when the awareness of women, who are much more likely to be victims, is not substantially affected. The results suggest that individual behavior can be affected by a change in the beliefs of other individuals, complementing experiments demonstrating that second-order beliefs can affect

---

slight uptick in state laws related to sexual misconduct, they were mostly limited in scope. Kelly, Cara, and Hegarty, Aaron - #MeToo was a culture shock. But changing laws will take more than a year. USA Today. October 4, 2018.

<sup>78</sup>The increase persists in the next wave of the survey conducted in November 2018-January 2019.

<sup>79</sup>A similar tipping point may have occurred around 2002 for clergy sexual abuse scandals (Bottan and Perez-Truglia, 2015).



behavior (Bursztyn et al., 2018). Second-order beliefs may have affected reporting if they changed victims' expectations regarding the response to reporting a sexual crime, either of the police or the wider community. In columns (3) and (4), we show that while awareness increased, agreement with the statement "*women who complain about harassment often cause more problems than they solve*" did not change substantially between 2016 and 2018. This suggests that reporting can increase even when the average stereotype associated with reporting does not substantially change.

## 6 Conclusions

This study shows that the MeToo movement had a substantial, persistent effect on the propensity to report sexual crimes to the police. This result is consistent across multiple samples and is robust across multiple estimation techniques. Focusing on the US allows us to analyze who was affected by the movement. The effect is strong and statistically significant for both sexual harassment and sexual assault. While the movement may have disproportionately focused on the experiences of white women of high socioeconomic status, it increased the reporting of sexual crimes to the police for both white and black victims, offenders, and counties, as well as in counties with both high and low socioeconomic status.

The heterogeneity results provide additional evidence for the causal effect of the MeToo movement, in contrast to some other event that occurred around October 2017. The MeToo movement focused on female victims and often on cases that occurred several months or years before they were discussed in the media. We find a strong significant effect among female victims and an especially strong effect among crimes that are reported at least a month after they occurred.

We estimate that the MeToo movement increased the number of sexual crimes reported by 25,870 in the first six months after the movement started. In the first 15 months, 66,658 crimes were reported as a result of the movement. Out of these crimes, 33,542 were sexual assaults reported in the US, and we find that the movement led to 4,174 arrests in the US.<sup>80</sup> The effect found in the US is equivalent to closing 25% of the gap between the reporting of sexual crimes and other violent

---

<sup>80</sup>We use the difference-in-difference specification to estimate an effect of the MeToo movement for each country separately and compare the actual number of reported sexual crimes with the predicted number of reported sexual crimes if the MeToo movement had not taken place. The calculation for the countries where we have partial police data (the US, the UK, and Australia) is based on the assumption that the MeToo movement had the same effect per-capita on areas for which we obtained data as in other areas in the country.

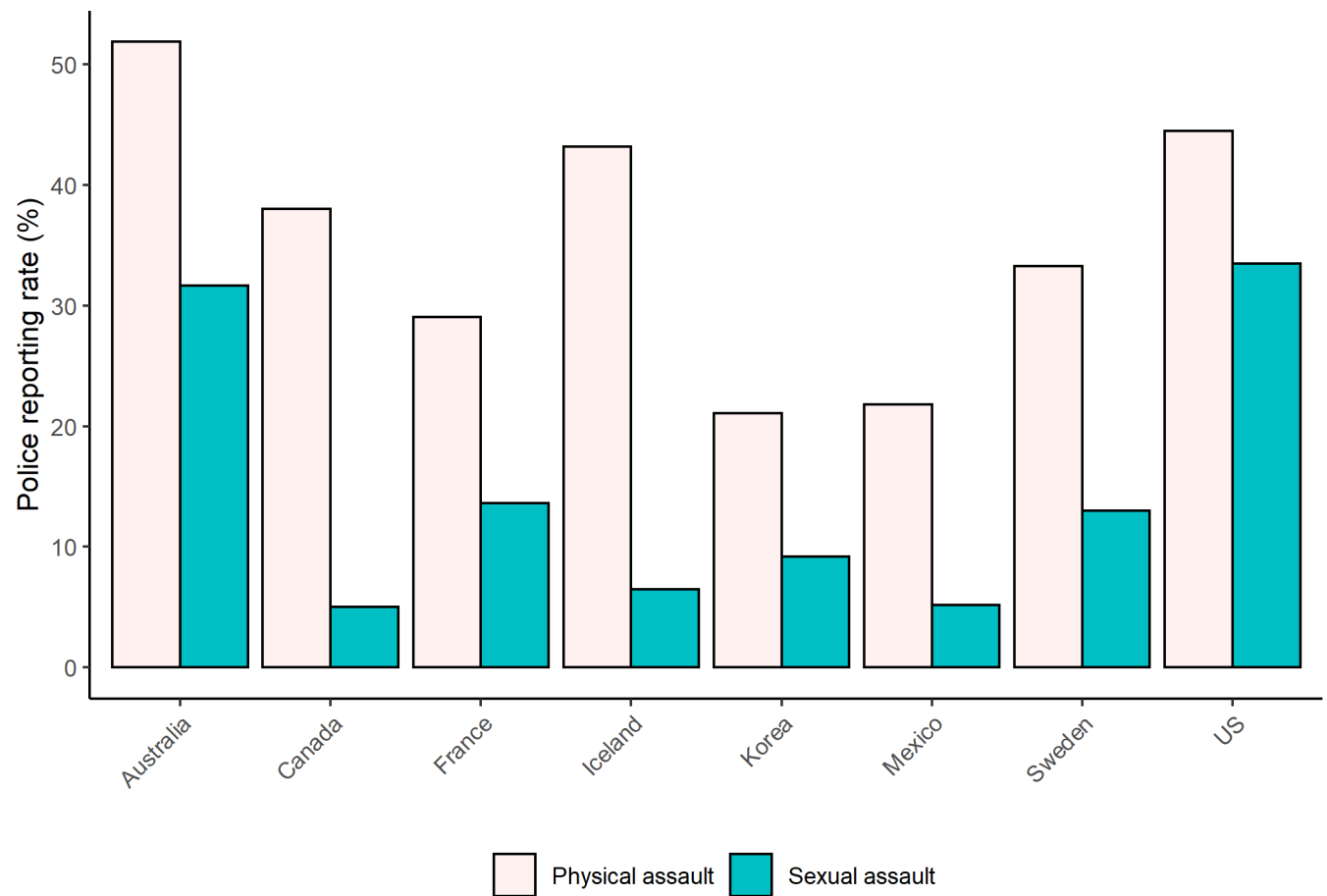
crimes observed in the National Crime Victimization Survey.

While the effect on the number of arrests is smaller than the effect on the number of reports, it is still an important channel through which reporting can have positive externalities. Increased arrests may deter offenders from committing future crimes, and if the arrests lead to convictions, they may also decrease the number of sexual crimes further by preventing potential repeat offenders from committing more crimes. Furthermore, even when a report does not lead to an arrest, it may lead to other disciplinary action, for example in a workplace or a university.

One limitation of this study is that it is difficult to disentangle the effect of the MeToo movement on the incidence of crimes from its effect on the propensity to report crimes. We show that a change in incidence cannot explain the effect found and is unlikely to drive the results. However, if the incidence of sexual crimes decreased as a result of the movement, our primary estimates should be interpreted as lower bounds for the increase in the propensity to report sexual crimes, as they are reduced by a lower incidence of crime.

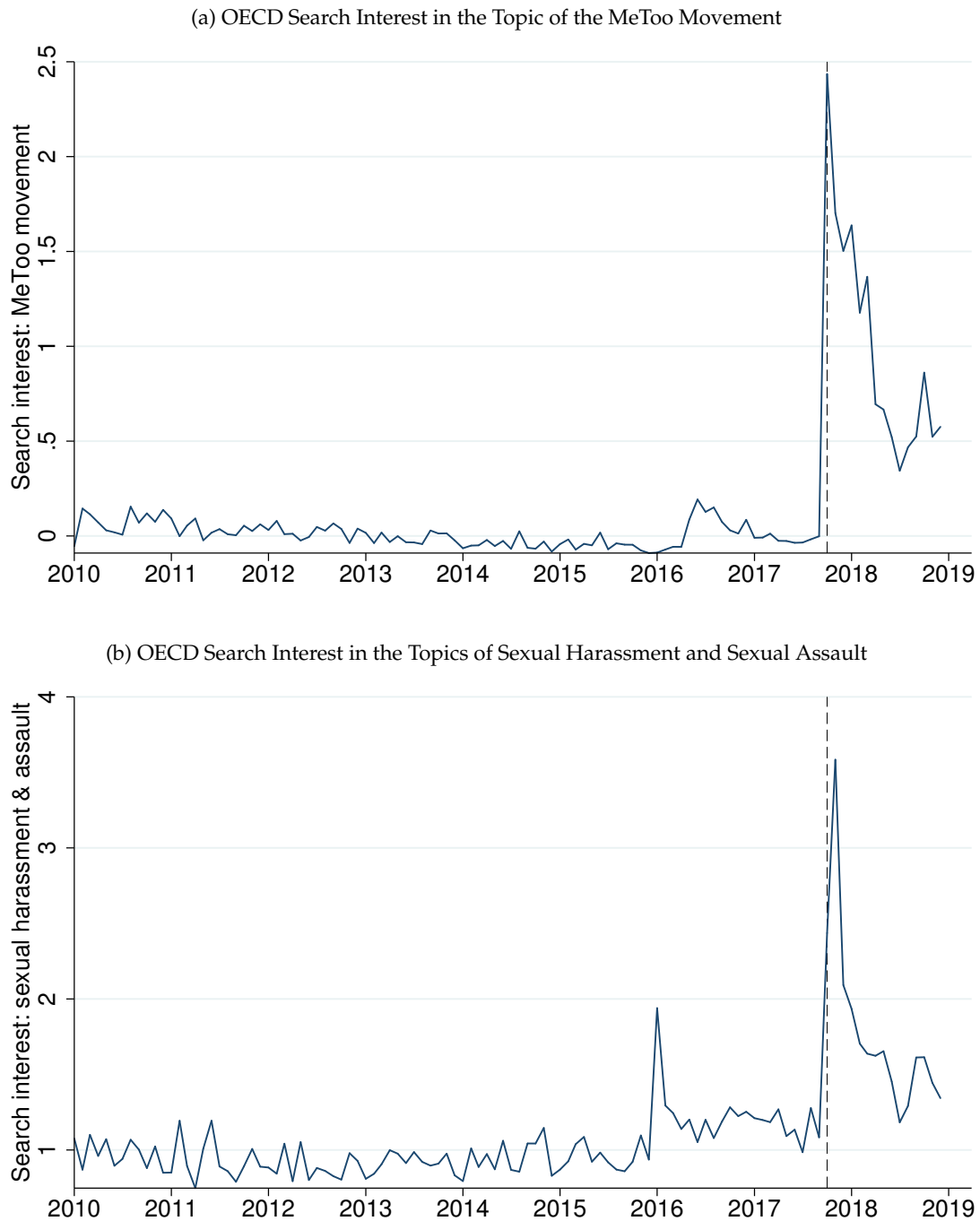
The findings show that social movements can have large, long-lasting effects on social norms, influencing individuals to make meaningful changes in their personal decisions. This effect may occur almost immediately and can change high stakes individual action.

Figure II.1: Share of Assaults Reported to the Police



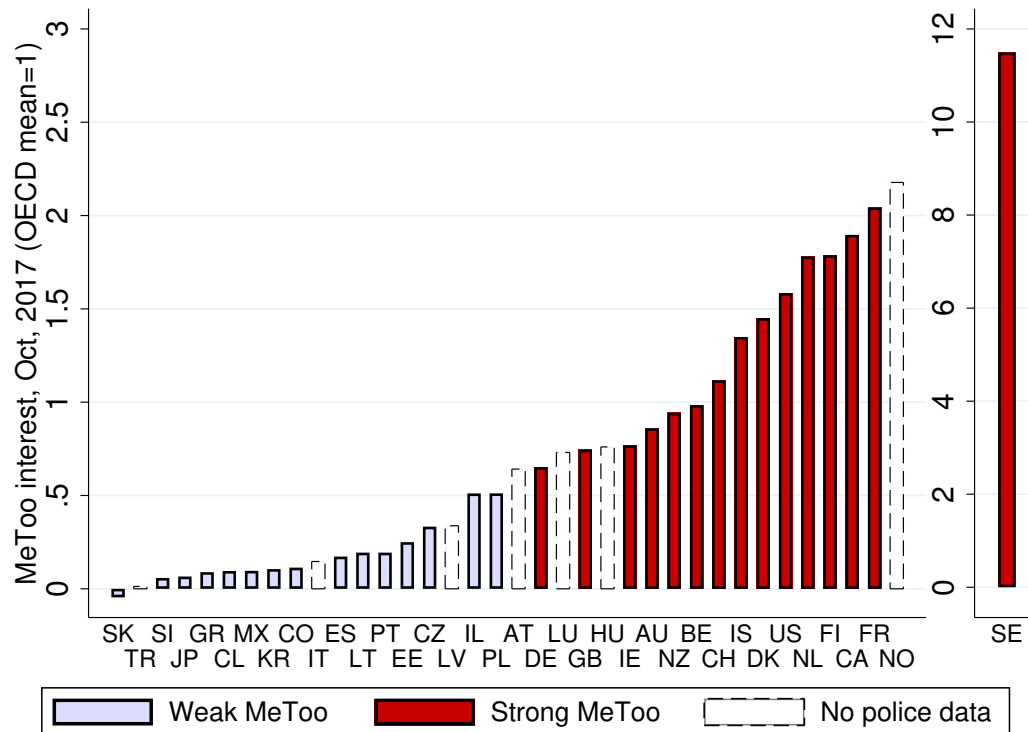
This figure shows the average share of sexual assaults and other physical assaults reported to the police in the years 2010-2017. Data is based on the UN Sustainable Development Goals, Indicator 16.3.1.

Figure II.2: Google Search Interest in the OECD



The figures show the monthly time series for the OECD means of both measures of the strength of the MeToo movement from 2010 to 2018. Data is from Google Trends. The vertical dashed line represents the start of the MeToo movement (October 2017). Sub-figure (a) shows search interest in the topic of the MeToo movement. Mean pre-MeToo interest is subtracted from the time series for each country separately so that the pre-MeToo period has a mean of zero, the variable is then normalized so that the post-MeToo OECD mean equals 1. Sub-figure (b) shows search interest in the topics of sexual harassment and sexual assault. The variable is normalized so that the pre-MeToo mean equals 1 for each country.

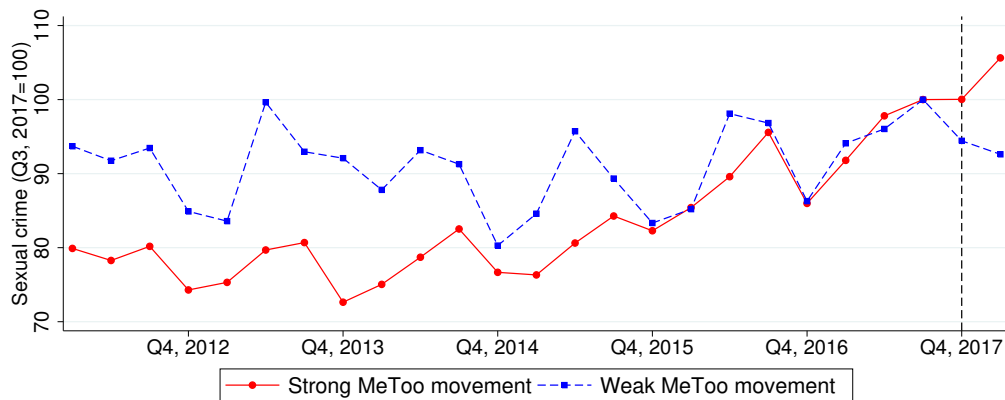
Figure II.3: Immediate Search Interest in the MeToo Movement



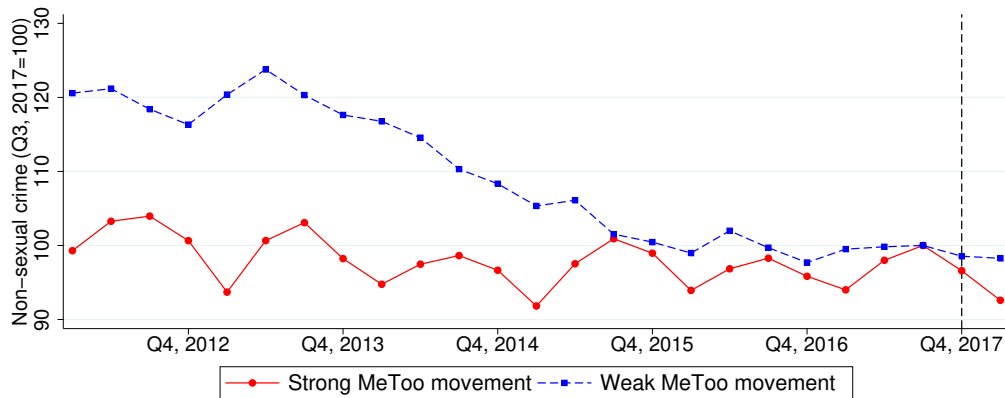
This figure shows the strength of the MeToo movement in OECD countries based on Google Search interest in the topic of the MeToo movement during October 2017. The Weak MeToo group of countries have below-median interest, the Strong MeToo group of countries have above-median interest, and the rest of the countries are not included in our sample since we have not obtained access to their police data.

Figure II.4: Crimes Reported over Time

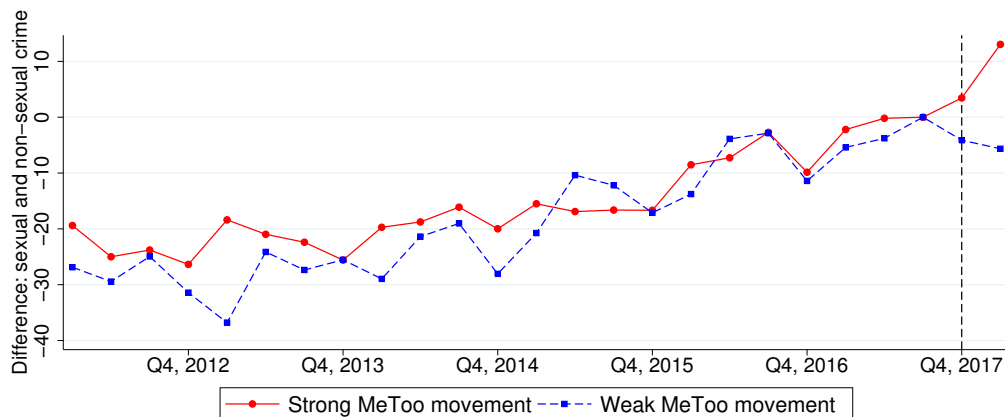
(a) Sexual Crime Reported in Countries with Strong and Weak MeToo Movements



(b) Non-Sexual Crime Reported in Countries with Strong and Weak MeToo Movements

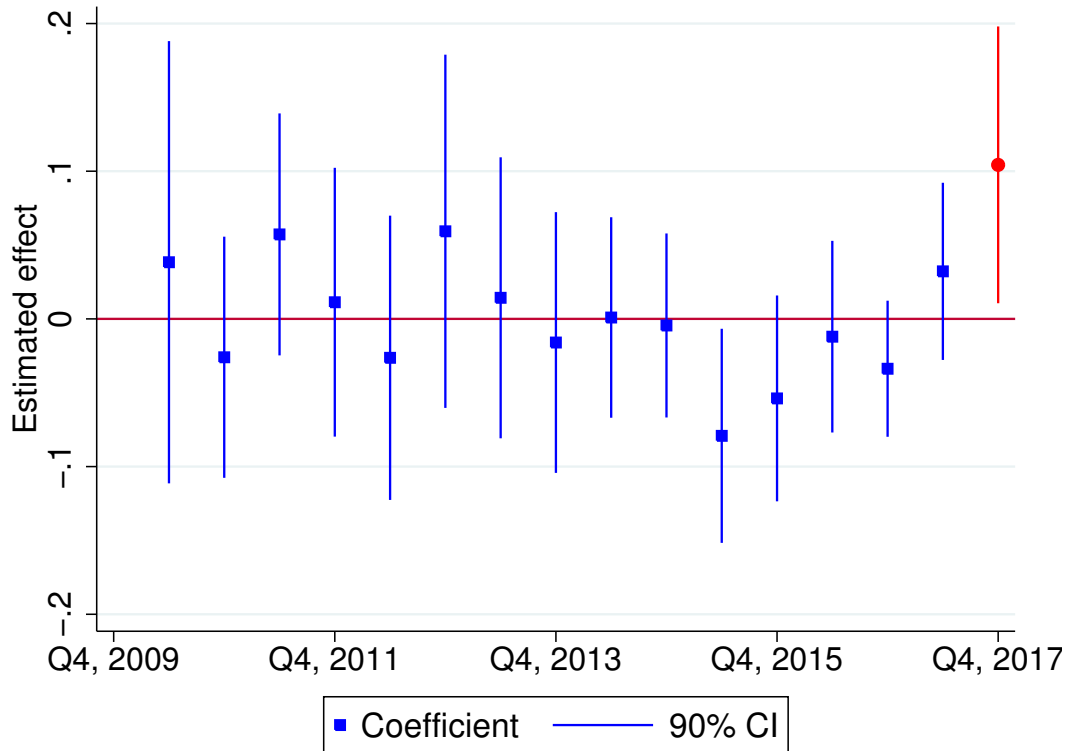


(c) Difference Between Sexual and Non-Sexual Crime Reported in Countries with Strong and Weak MeToo Movements



Figures (a) and (b) show the number of reported sexual crimes and the number of reported non-sexual crimes, both normalized to 100 in Q3 2017 for each country, and averaged separately for the countries with strong and weak MeToo movements. Figure (c) shows the difference between the normalized number of sexual crimes and the normalized number of non-sexual crimes. The vertical dashed line represents the start of the MeToo movement, Q4, 2017. Data include all 30 countries in our sample. For four countries data is available for only part of the period, see Appendix Table II.A.5 for details.

Figure II.5: Placebo Tests, Setting the Start Date of the MeToo Movement in Every Second Quarter from Q2 2010 to Q4 2017



This figure shows the results from 15 placebo triple-difference regressions (Q2 2010-Q2 2017) and our main triple-difference result (Q4 2017, shown in red). Each coefficient comes from a regression using the full Q1 2010 - Q1 2018 dataset, but with a different six-month period for when the placebo MeToo movement happened. The corresponding confidence intervals are constructed using standard errors clustered at the country by crime type level.

Table II.1: Effect of the MeToo Movement During the First Six Months

	ln(crime)			
	(1)	(2)	(3)	(4)
Post * Strong MeToo	0.114** (0.048)		0.009 (0.031)	0.009 (0.031)
Post * Sexual crime		0.072** (0.030)		0.019 (0.044)
Post * Strong MeToo * Sexual crime			0.123*** (0.036)	0.104* (0.057)
Post * Weak MeToo * Sexual crime			0.019 (0.044)	
Country * Crime type * Lin. trend	X	X	X	X
Country * Crime type * Quarter	X	X	X	X
Post	X	X	X	X
Crime data used	Sexual crimes	All crimes	All crimes	All crimes
Final quarter	Q1 2018	Q1 2018	Q1 2018	Q1 2018
Observations	904	1,808	1,808	1,808
Clusters	30	60	60	60

This table shows the effect of the MeToo movement on sexual crimes reported using data from 30 OECD countries for the period from Q1 2010 to Q1 2018. Column (1) uses data on sexual crime only while Columns (2)-(4) uses data on both sexual and non-sexual crimes. A country is categorized as having a strong MeToo movement if search interest for the topic of the MeToo movement was above the OECD median in October 2017. Standard errors clustered at the country by crime level in parenthesis. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .



Table II.2: Persistence of the Effect in Countries with a Strong MeToo Movement

	ln(crime)	
	(1)	(2)
Post * Sexual crime	0.104*** (0.035)	
2017 Q4 * Sexual crime		0.121*** (0.033)
2018 Q1 * Sexual crime		0.122** (0.051)
2018 Q2 * Sexual crime		0.083** (0.037)
2018 Q3 * Sexual crime		0.087** (0.037)
2018 Q4 * Sexual crime		0.108** (0.043)
Country * Crime type * Lin. trend	X	X
Country * Crime type * Quarter	X	X
Post	X	
Q4 2017-Q4 2018 FE		X
Final quarter	Q4 2018	Q4 2018
Observations	1,012	1,012
Clusters	30	30

This table shows the effect of the MeToo movement over time using data from the 15 OECD countries with a strong MeToo movement in October 2017. Standard errors clustered at the country by crime level in parenthesis. \*\*\*p<0.01, \*\*p<0.05, \*p<0.1.

Table II.3: Robustness Checks

(1) Preferred specification	0.104* (0.057)
Length of short-term period:	
(2) 3 month effect	0.060 (0.063)
(3) 9 month effect	0.095* (0.055)
Different measures of MeToo strength:	
(4) 6m MeToo search interest	0.102* (0.060)
(5) SA/SH immediate search interest	-0.015 (0.061)
(6) % heard of MeToo movement	0.095 (0.080)
Alternative specifications:	
(7) Weighted by country population	0.119** (0.052)
(8) Only data based on date crimes were reported	0.119* (0.065)
(9) Outcome variable: Normalized number of crimes	0.112* (0.057)
(10) Negative binomial regression	0.118** (0.048)
Alternative empirical strategies:	
(11) Matrix completion method	0.171*** (0.043)
(12) 2SLS: Fraction Eng. speakers as IV	0.096 (0.071)

This table shows robustness checks for our main triple-difference estimate. Row (1) repeats the main estimate from Column (4) of Table II.1. Rows (2)-(3) use different periods over which the effect is measured. Rows (4)-(6) use different measures of the strength of the MeToo movement. Row (7) shows the result of our main specification weighted for the countries population. Row (8) only includes data from the 24 countries basing their statistics on the date the crimes were reported to the police. Row (9) uses the normalized number of crimes as the outcome variable. Crimes are normalized to be one on average for each country by crime type group, in the year leading up to the start of the MeToo movement. Row (10) shows the result of a negative binomial regression using the count data of crimes reported as the outcome variable. Row (11) shows the result of using the matrix completion method. Row (12) shows the result of a two-stage least squares regression where having a strong MeToo movement is instrumented for by the fraction of English speakers. All rows except Rows (6) and (8) use data from 30 OECD countries. Row (6) uses data from the 12 OECD countries surveyed in the 2019 YouGov survey. Standard errors clustered at the country by crime level are in parenthesis. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

Table II.4: Effect of the MeToo Movement on Sexual Crimes in the US

	ihs(crime)		
	(1)	(2)	(3)
Post * Sexual Assault	0.081*** (0.015)		
Post * Sexual Assault		0.112*** (0.036)	
Post * Sexual Harassment		0.148*** (0.055)	
Post * Sexual Crimes			0.129*** (0.036)
State * Crime Type * Lin. Trend	X		
State * Crime Type * Month	X		
City * Crime Type * Lin. Trend		X	X
City * Crime Type * Month		X	X
Post	X	X	X
Data	NIBRS	City	City
Final Month	Mar 2018	Mar 2018	Mar 2018
Observations	6,654	1,863	1,242

This table shows the effect of the MeToo movement on sexual crimes reported based on NIBRS and city crime data. Regressions are weighted by the number of crimes that occurred in each state/city before the MeToo movement started. Robust standard errors in parenthesis. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

Table II.5: Effect of the MeToo Movement by Relationship and Crime Type

	ihs(crime)		
	(1)	(2)	(3)
Post * Fondling	0.111*** (0.019)		
Post * Rape	0.093*** (0.017)		
Post * Sodomy	-0.024 (0.031)		
Post * Statutory Rape	0.027 (0.042)		
Post * Sexual Assault, No Injury		0.093*** (0.016)	
Post * Sexual Assault, Injury		0.028 (0.022)	
Post * Sexual Assault, Knew Offender			0.089*** (0.016)
Post * Sexual Assault, Stranger			0.104*** (0.035)
Difference		0.065***	-0.015
State * Crime Type * Lin. Trend	X	X	X
State * Crime Type * Month	X	X	X
Post	X	X	X
Final Month	Mar 18	Mar 18	Mar 18
Observations	16,635	9,981	9,981

This table shows the effect of the MeToo movement on different crime types. In each column, crimes are aggregated into different categories. The reference group for all columns is non-sexual crimes. In Column (1), the category "Sexual Assault With An Object" is excluded since approximately a third of state\*months had zero crimes reported. Incidents related to multiple sexual offense crime categories are also excluded. In Column (2), cases where it is unknown if a victim was injured are excluded. In Column (3), cases where the relationship between the victim and offender was not reported or where the relationship is unknown are excluded. 2010-2018 NIBRS data. Regressions are weighted by the number of crimes that occurred in each state before the MeToo movement started. Robust standard errors in parenthesis. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

Table II.6: Effect of the MeToo Movement by the Lag Between the Occurrence and Reporting Dates

	(1)	(2)
Post * Sexual Crimes, Lag<=30 Days	0.095** (0.038)	0.111*** (0.023)
Post * Sexual Crimes, Lag>30 Days	0.215*** (0.049)	0.135*** (0.048)
City * Crime Type * Lin. Trend	X	X
City * Crime Type * Month	X	X
Post	X	X
Treatment Dates	Oct 17-Mar 18	Apr 18-Dec 18
Pre period mean Lag<=30	170	170
Pre period mean Lag>30	44	44
Observations	1,842	1,905

This table shows the effect of the MeToo movement on sexual crimes according to when the crime was reported. In all columns, the data is aggregated into three categories: sexual crimes reported within 30 days, sexual crimes reported after more than 30 days, and non-sexual crimes. Non-sexual crimes is the reference category. Column (1) focuses on the primary main short-term effect and includes data until March 2018 and Column (2) excludes October 2017-March 2018. Regressions are weighted by the number of crimes that occurred in each city before the MeToo movement started. City crime data 2010-2018. Robust standard errors in parenthesis. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

Table II.7: Effect of the MeToo Movement by Victim and Offender Demographics

	lhs(crime)			
	(1)	(2)	(3)	(4)
Post * Sexual Assault, Victim Female	0.091*** (0.016)			
Post * Sexual Assault, Victim Male	0.033 (0.024)			
Post * Sexual Assault, Victim Black		0.077*** (0.024)		
Post * Sexual Assault, Victim White		0.082*** (0.016)		
Post * Sexual Assault, Offender Female			0.015 (0.042)	
Post * Sexual Assault, Offender Male			0.098*** (0.016)	
Post * Sexual Assault, Offender Black				0.095*** (0.022)
Post * Sexual Assault, Offender White				0.092*** (0.017)
Difference	0.058**	-0.005	-0.083*	0.003
State * Crime Type * Lin. Trend	X	X	X	X
State * Crime Type * Month	X	X	X	X
Post	X	X	X	X
Final Month	Mar 18	Mar 18	Mar 18	Mar 18
Observations	9,981	9,981	9,981	9,981

This table shows the effect of the MeToo movement by victim and offender demographics. In each column, crimes are aggregated into different categories. The reference group for all columns is all non-sexual crimes. In Columns (1) and (3), crimes where the sex of the victim or the offender is unknown are excluded along with crimes with multiple victims or offenders. In Columns (2) and (4), crimes with a single white or black victim or offender are included. 2010-2018 NIBRS data. All regressions are weighted by the number of crimes that occurred in each state before the MeToo movement started. Robust standard errors in parenthesis. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

Table II.8: Effect of the MeToo Movement by County Demographics

	ihts(crime)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post * Sexual Assault	0.088*** (0.011)	0.088*** (0.011)	0.088*** (0.011)	0.088*** (0.011)	0.088*** (0.011)	0.088*** (0.011)	0.088*** (0.011)
Post * Sexual Assault * Med. Income (std. dev.)		0.013 (0.009)					
Post * Sexual Assault * % College			0.127 (0.098)				
Post * Sexual Assault * % Blacks (Compared to Whites)				0.071 (0.075)			
Post * Sexual Assault * % Other Race (Compared to Whites)					0.557*** (0.178)		
Post * Sexual Assault * % Hispanics						0.309*** (0.111)	
Post * Sexual Assault * % Vote Trump							-0.266*** (0.071)
Interquartile Range of Demographic Diff. in Effect * 75th-25th Pct.		1.207 0.016	0.132 0.017	0.194 0.014	0.054 0.03	0.062 0.019	0.265 -0.071
County * Crime Type * Lin. Trend	X	X	X	X	X	X	X
County * Crime Type * Month	X	X	X	X	X	X	X
Post	X	X	X	X	X	X	X
Post * Demographic	X	X	X	X	X	X	X
Final Month	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18
Observations	170,564	170,564	170,564	170,564	170,564	170,564	170,564

This table shows the effect of the MeToo movement based on county-level data and tests for heterogeneous effects by county demographics. 2010-2018 NIBRS data. All regressions are weighted by the number of crimes that occurred in each county before the MeToo movement started. All demographic variables are first subtracted by their weighted mean. Robust standard errors in parenthesis. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

Table II.9: Effect of the MeToo Movement on Arrests

	ihts(crime)					
	(1)	(2)	(3)	(4)	(5)	(6)
Post * Sexual Assault, No Arrest	0.095*** (0.016)			0.105*** (0.011)		
Post * Sexual Assault, Arrest	-0.008 (0.026)			0.052*** (0.018)		
Post * Sexual Assault		0.014 (0.027)	0.091*** (0.016)		0.071*** (0.019)	0.107*** (0.011)
Difference	0.103***			0.053***		
State * Crime Type * Lin. Trend	X	X	X	X	X	X
State * Crime Type * Month	X	X	X	X	X	X
Post	X	X	X	X	X	X
Final Month	Mar 18	Mar 18	Mar 18	Dec 18	Dec 18	Dec 18
Crimes	All	Arrest	No Arrest	All	Arrest	No Arrest
Observations	9,981	6,654	6,654	10,899	7,266	7,266

This table shows the effect of the MeToo movement on sexual crimes by whether an arrest was made. A case is defined to have an arrest if a suspect is taken into custody based on a warrant or previously submitted report, arrested on view without a warrant or summoned to court. In Column (1) and (4), the crimes are aggregated to three separate crime categories: sexual crimes where an arrest was made, sexual crimes where no arrest was made, and non-sexual crimes, which are the control group. In Column (2) and (5), only crimes where an arrest was made are included and columns (3) and (6) include only crimes where no arrest was made. Columns (1)-(3) focus on the short-run effect and columns (4)-(6) focus on the long-run effect. 2010-2018 NIBRS data. Regressions are weighted by the number of crimes that occurred in each state before the MeToo movement started. Robust standard errors in parenthesis. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1



Table II.10: Effect on Crimes that Occurred Before the MeToo Movement Started

	lhs(crime)
Post * Sexual Crimes	0.194** (0.077)
City * Crime Type * Lin. Trend	X
City * Crime Type * Month	X
Post	X
Final Month	Dec 2017
Crimes Included	3 Month <= Lag
Observations	1,179

This table shows the effect of the MeToo movement on sexual crimes, which were reported at least three months after they occurred. The table only includes crimes reported by December 2017. Therefore, all crimes included in this table occurred before the MeToo movement started. 2010-2017 city crime data. Regressions are weighted by the number of crimes in each city before the MeToo movement started. Robust standard errors in parenthesis. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

Table II.11: Change in Beliefs Regarding Sexual Harassment

	Workplace sexual harassment no longer a problem		Accusers cause more problem than they solve	
	(1)	(2)	(3)	(4)
April-May 2018	-0.136*** (0.032)		-0.010 (0.025)	
Women, 2018		-0.047 (0.042)		0.004 (0.034)
Men, 2018		-0.234*** (0.047)		-0.026 (0.035)
Ref. Group	2016	2016	2016	2016
Respondent FE	X	X	X	X
Observations	9,252	9,236	9,212	9,196

This table shows the change in beliefs regarding sexual harassment between 2016-2018. The data is the pooled 2016 and 2018 responses for the Views of the Electorate Research Survey. Columns (1) and (2) refer to respondents' agreement with: "Sexual harassment against women in the workplace is no longer a problem in the United States." Columns (3) and (4) refer to respondents' agreement with "Women who complain about harassment often cause more problems than they solve." The answers are coded between 0 (strongly disagree) and 3 (strongly agree) and then standardized. The results are similar when a binary coding of the response is used instead. All regressions control for respondent fixed effects. Robust standard error in parenthesis. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

# Appendices

## A Data Processing

### A.1 Crime Classification

For both the US and international data we classify each crime as belonging to one of the following categories: sexual assault, defined as a sexual crime that includes physical contact; sexual harassment, defined as a sexual crime that does not include physical contact (e.g. stalking or indecent exposure); non-sexual crimes and crimes which are not directly affected by the MeToo movement but could be indirectly related to it. Crimes indirectly related to the MeToo movement include crimes related to bestiality, bigamy, crime against children, domestic assault, harassment where it is not clear if the harassment is of sexual nature, incest, pedophilia, pornography, prostitution, and registration of sexual offenders. We exclude these crimes from the analysis since spillovers from the MeToo movement can affect this group of crimes, and therefore, they are not a suitable control group. We also exclude from the analysis cases appearing in police records that are not related to any specific crime (e.g., missing person investigation) and traffic tickets.

Throughout most of the analysis, we aggregate the sexual assault and sexual harassment crimes into one category, defined as sexual crime.

### A.2 OECD Crime Data Collection and Processing

To collect high-frequency crime data from as many OECD countries as possible, we first downloaded the data available on the websites of the statistics agencies and the police. If no data was available online, we contacted both the main statistics agency as well as the national police requesting data on the number of crimes reported at a monthly or quarterly level. Finally, if these contacts did not yield the required data, we filed the equivalent of a Freedom of Information Act request or purchased data specifically aggregated for our project from the statistics agency.

To quality control our international data, we crosschecked our data for the 19 EU countries in our sample with the 2017 Eurostat data on sexual violence. To avoid correlation between the two datasets driven by the population size of the countries, we compared the sexual crimes per

population of 100,000. Reassuringly, the correlation in the number of sexual crimes per population of 100,000 is 0.96. The average percentage difference between the numbers in the two datasets is -1% showing that there is no systematic difference in the level of the numbers and corroborating that the data we collected is in line with EU estimates. Finally, the average absolute percentage difference between the numbers in the two datasets is 24% showing that for most estimates the two numbers are similar in magnitude. The difference could be explained by the fact that we excluded specific sexual crimes that did not seem directly related to the MeToo movement (such as crimes against children) and since we include crimes that can appear outside the sexual assault category, such as stalking.

In Australia, the United Kingdom, and the United States, high-frequency data on the number of crimes reported are not available for the whole country.<sup>81</sup> For Australia, we have data for New South Wales, Queensland, Victoria, and Western Australia, covering 88% of the population, but not for the Australian Capital Territory, Northern Territory, South Australia, and Tasmania. For the United Kingdom, we have data for England, Northern Ireland, and Wales, covering 92% of the population, but not Scotland. For the United States, we use the NIBRS data described in more detail in Appendix Section A.5.

The 30 countries in our dataset are listed in Appendix Table II.A.5 together with the organizations providing the data, the time period covered as well as the percentage of the population covered by the police agencies providing the data.

For most countries in the data, the quarter that a crime is counted in is based on the date the crime was reported. For four countries, Belgium, Colombia, Germany, and Iceland crimes are counted in the quarter when they occurred. For the UK and US, some of the crimes are counted in the quarter they were reported while other crimes are counted in the quarter they occurred. For Switzerland, the crime is counted in the quarter information about the case was transmitted to the Federal Statistical Office, which for the vast majority of crimes is in the same quarter as the crime is reported to the police. Only one of the countries not providing data based on the date the crimes were reported is a country classified as having a weak MeToo movement. Therefore, the small

---

<sup>81</sup>In the US, crime data for many agencies is also available through the UCR Summary Report System. We do not use that dataset since the definition for rape has changed in 2013 and agencies are gradually changing their reports based on the new definition. Furthermore, this system only collects data on the most severe crimes and therefore it does not include data on sexual assaults besides rapes.

effect of the MeToo movement in countries with weak movements cannot be explained by the data from these countries being based on the date of occurrence as opposed to the date of reporting.

### A.3 Google Search Data Processing

As our primary measure of the MeToo movement's strength, we use the search interest in the topic of the MeToo movement in October 2017. Our search interest data is scraped from Google trends and contains monthly search interest figures for all of the OECD from 2010-2018.<sup>82</sup> To ensure that our primary measure of MeToo movement strength is not higher for countries that more frequently use search terms related to the MeToo movement, before these terms had been given the meaning they were given by the MeToo movement, we difference out the average search intensity for these terms from the period before the MeToo movement for each country, so that each country has an average interest of zero in the pre-period. Finally, to simplify the interpretation of this measure, we normalize the magnitude of the interest so that the average interest in the OECD is one in the post-period.

Google does not provide information on the phrases defined as being part of the MeToo movement topic. Therefore, we also create our own definition of the MeToo movement topic in all of the languages used in the OECD, for which we could find a phrase related to the MeToo movement. We restricted our measure to phrases with search interest in their country of origin of at least 1% of the search interest for "me too" in the US, these terms are: "me too", "balance ton porc", "moi aussi", "quella volta che" and "yo tambien" as well as these terms written without spaces.<sup>83</sup> In October 2017, searches for these phrases has a 0.997 correlation with the MeToo movement topic defined by Google across countries. We prefer to use the search for the MeToo movement topic instead of our list of exact phrases since it is more likely that the topic search will include searches for additional phrases related to the MeToo movement in other languages.

In Tables II.A.1 and II.3, we use an alternative measure of search interest based on searches related to the topics of sexual harassment and sexual assault. Again, the topics are defined by Google as all searches that include the concept of sexual harassment or sexual assault in any

---

<sup>82</sup>For scraping, we used the R package gtrendsR written by Philippe Massicotte and Dirk Eddelbuettel. The data was scraped on October 26, 2020.

<sup>83</sup>We exclude searches that contained the term "me too" along with the words "meghan", "trainor" or "song" since the song "Me too" by Meghan Trainor caused an increase in search interest around its release in May 2016.

language. In contrast to searches for the MeToo topic, searches for the topics of sexual harassment and sexual assault have the same interpretation before and after the start of the MeToo movement. Therefore, we normalize the search interest so that the pre-MeToo period mean is one for each country.

#### **A.4 Fraction of English Speakers Data Processing**

We use data on the fraction of English speakers from the 23rd edition of the Ethnologue Global Dataset. The data contains estimates for the population using English as their first language, the population using English but for whom English is not a native language and the total population. We divide the population using English as a first language by the population to get the fraction of first-language English users. We take the sum of the population using English as a first language and the population using English as a non-native language and divide it by the total population to get the fraction of the population who uses English.

For five countries (Chile, Colombia, France, Slovenia, and Slovakia) we do not have an estimate for the number of first-language English users. We impute the fraction using English as their first language using the median fraction of first-language English users for the country's region (South America, Western Europe, Southern Europe, and Eastern Europe). For Japan, there is no estimate for the fraction of non-native English users in the Ethnologue data. Instead, we use an estimate of 5% provided in communications with Ethnologue and confirmed in a report from Mitsue-Links.<sup>84</sup>

#### **A.5 NIBRS Crime Data Processing**

We classify NIBRS offenses as either sexual assault or non-sexual crimes. The sexual assault offenses are fondling, rape, sexual assault with an object, sodomy, and statutory rape. We exclude incest, human trafficking, and the pornography/obscene material crime categories. All other 43 offense types form the non-sexual crimes category. Domestic assault is not a separate offense type in the NIBRS dataset. To exclude domestic violence crimes which may have been affected by the MeToo movement, we exclude all aggravated assaults where the circumstances of the assault are defined in the NIBRS as a "lovers quarrel" and all assaults or aggravated assaults for which the relationship

---

<sup>84</sup>[https://www.mitsue.co.jp/english/global\\_ux/blog/201709/14\\_1700.html](https://www.mitsue.co.jp/english/global_ux/blog/201709/14_1700.html)

between the offender and victim is defined in the NIBRS as one of the following: victim was ex-spouse, victim was spouse, homosexual relationship, victim was boyfriend/girlfriend, victim was common-law spouse.

In the NIBRS data, an incident can include multiple crimes if they occurred in concert, at the same time and place. Since our classification of incidents depends on the type of offense committed, we define an incident as a sexual assault if one of the offenses which occurred as part of the incident is a sexual assault. Similarly, if the incident is not a sexual assault, we exclude it if one of the offenses which occurred as part of the incident should be excluded (e.g., if an incident includes both a pornography/obscene material offense and a weapon law violations offense, it will be excluded).

When analyzing state-level data, we exclude state-years where there are months with fewer than 100 crimes reported in total.

One potential concern is that police agencies started reporting sexual assaults through the NIBRS as a result of the MeToo movement. However, we find no evidence that the movement affected reporting or that agencies determine when to include sexual assault in their reports. We check whether agencies participating in the NIBRS system started reporting sexual assaults in a specific month. Since there is natural variation in reporting, we focus on cases where agencies did not report any sexual assaults in twelve consecutive months and then reported at least four assaults. This occurred in only seven agencies out of over 2,000. Even in those agencies, the increase is from zero cases reported to four or five cases, and the increase does not occur after the MeToo movement started. Therefore, this increase probably represents noise and not a decision of an agency to start reporting sexual assaults.

## **A.6 City Crime Data Processing**

Data for each city was obtained separately from the city's open data website. For each city, we first categorize a crime as a sexual assault, sexual harassment, non-sexual crime, or a crime that should be excluded since it is indirectly related to the MeToo movement (as explained in Appendix Section A.1). If an observation is defined at the crime level and the data include multiple crimes per incident, we then aggregate crimes at the incident level. The incident crime category is defined as the most severe crime of the crimes composing the incident, where we use the following hierarchy:

Sexual assault, sexual harassment, excluded crimes, other crimes.<sup>85</sup>

In the city data, we define each month as spanning from the 15th day of the calendar month to the 14th day of the next calendar month. By defining months in this way, we can cleanly categorize each observation in the aggregated data as occurring before or after the start of the MeToo movement, since the movement started on October 15, 2017.<sup>86</sup>

## B Additional Analysis

### B.1 Allowing for Different MeToo Start Dates in Each Country

In addition to the strategy described in Section 3.4, we also use the variation in the start dates of the movement to estimate the effect of the movement over time. We restrict the sample of countries to countries that at some time before the end of 2018 had a MeToo movement and use the following regression:

$$y_{itc} = \beta_1 MeToo_{ct} \times SexCrime_i + \beta_2 MeToo_{ct} + \beta_{3,ic} Trend_t + \gamma_{i,c,q(t)} + \varepsilon_{itc} \quad (11)$$

where  $MeToo_{ct}$  takes a value of one if the MeToo movement in country  $c$  started before or in the first month of quarter  $t$  and  $MeToo_{ct}$  equals one-third or two-thirds if the movement started in the second or third month of quarter  $t$ , respectively. The other terms of the equation are defined in the same way as in Equation 8.

We use two different strategies for estimating the start of the MeToo movement in each country. First, we define the start of the movement as the first month when Google search interest in the MeToo topic was higher than the OECD median in October 2017. Under this classification, all the countries that were classified as having had strong MeToo movements in the analysis in Section 3.3 have MeToo movements starting in October 2017, but additional countries have MeToo

---

<sup>85</sup>Typically, multiple crimes which form an incident occur at the same date. However, in Kansas City, an incident (or a “case”) can be continuously updated and appear multiple times in the dataset, for example, when the victim reports a crime and when the police has a suspect. In cases where an incident appears more than once in the dataset and includes at least one report from a victim, we include only the report of the victim. If an incident still has multiple updates, we include only one observation and define the date the incident was reported as the minimal date among all observations related to the incident. If an incident is associated with crimes that occurred over multiple days, we define the date the incident occurred as NA.

<sup>86</sup>We do not use a similar definition when analyzing the international data or NIBRS data since most international data we collect is already aggregated at the month or quarter level, and since we want to keep the NIBRS results consistent with the international analysis.



movements starting after October 2017. Our second criterion is based on search interest for the sexual harassment and sexual assault topics. We classify the start of a MeToo movement as the first month, in or after October 2017, that had the highest search interest for the sexual harassment and sexual assault topics since 2010. Appendix Table II.A.2 shows the start dates of the MeToo movement for each country according to both criteria. Note that both these specifications have potential reverse causality problems since an increase in sexual crimes reported to the police could affect searches for the topics of the MeToo movement, sexual harassment, and sexual assault. Due to this problem, we see this analysis as supplemental to our main analysis.

Table II.A.1 shows the results of our analysis using different start dates of the MeToo movement. Columns (1) and (2) use the start date based on searches for the MeToo topic, and Columns (3) and (4) use the start date based on sexual harassment and sexual assault topic searches. Column (1) reports an overall effect of the MeToo movement of 12%. Column (2) splits these estimates by the number of quarters since the start of the MeToo movement and shows that the effect is stable over time with all of the point estimates for each of the quarters since the start of the MeToo movement being between 9% and 13%.<sup>87</sup> The results of Column (3)-(4) are slightly smaller but qualitatively similar.

## B.2 Neighborhood-Level Heterogeneity

In this section, we analyze heterogeneous effect of the MeToo movement by neighborhood demographics in the sample of seven large US cities. To determine the neighborhood where each crime occurs, we use the most coarse definition of police administrative areas available in the dataset. We use the most coarse definition (e.g., a police division instead of a police beat) to ensure that the number of crimes is positive for most observations. The jurisdictions are detailed in Appendix Table II.A.3. In the case of Nashville, the police precinct where the crime occurred is not reported in the city's crime dataset, and we identify the precinct based on the rounded coordinates of the crime's location.

We use the shapefiles for the police boundaries of each city to identify the geographical boundaries of each neighborhood. For most cities, we use the most recent shapefile available. For

---

<sup>87</sup>In Columns (2) and (4), the dummy variables take the value of either one or zero, regardless of when in a quarter the MeToo movement started.

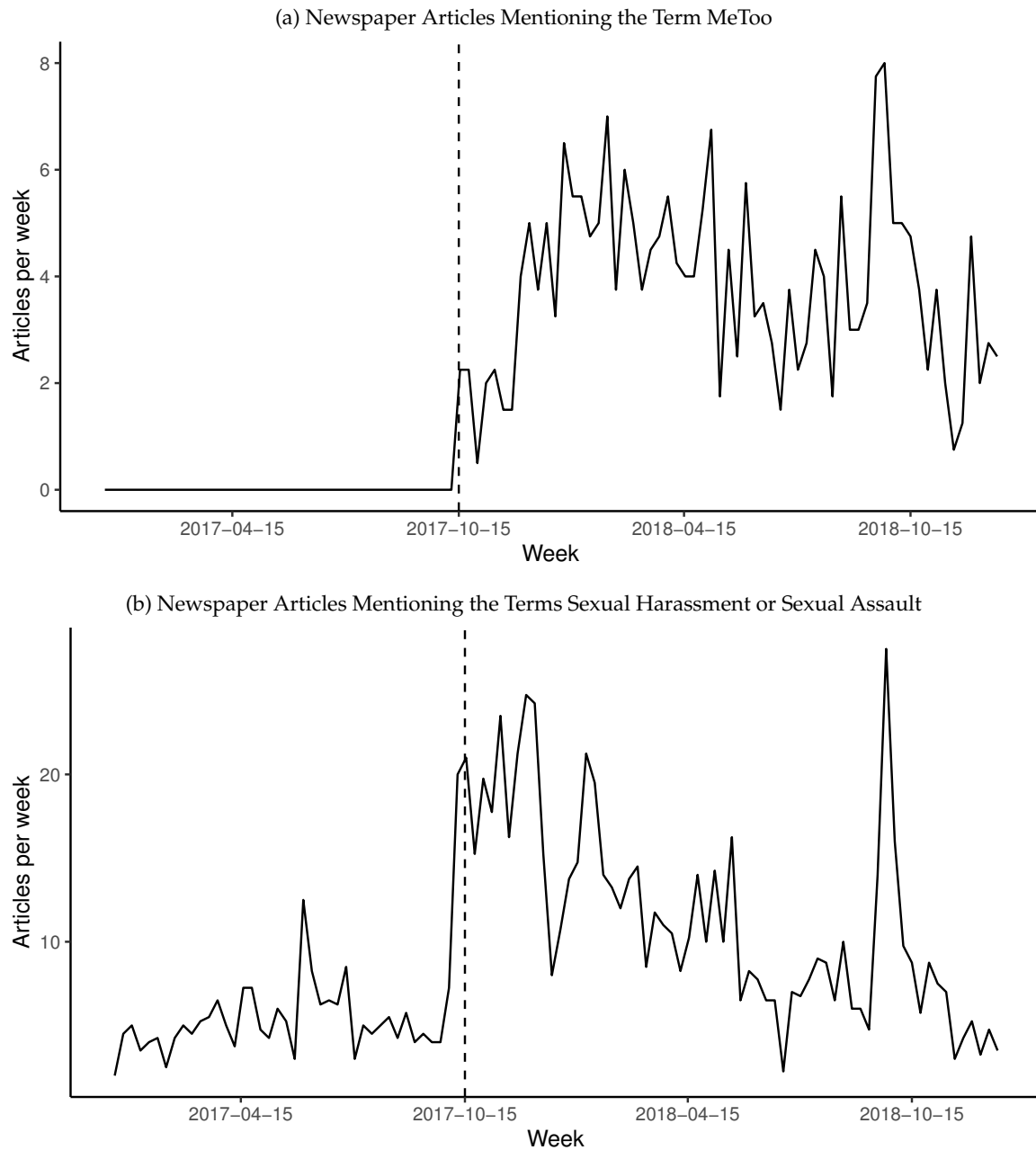
Seattle, where changes in the shapefiles are clearly defined, we use different shapefiles for different years and determine the boundaries of each police precinct according to the year when the crime occurred.

The demographics of each neighborhood are determined by spatially matching the neighborhood with census block groups. We calculate each neighborhood's demographics as the weighted average of the demographic covariates among overlapping block groups, where the weight of each block group is the population of the block group multiplied by the share of the block group's area overlapping with the neighborhood. The demographics for each block group are based on the American Community Survey 5-year 2016 estimates.

Table II.A.4 does not find evidence for strong heterogeneity by the neighborhood demographics. While some of the point estimates are consistent with a stronger movement among higher-income and college-educated neighborhoods, the estimated heterogeneity is relatively small. For example, the difference in the expected effect on reporting between a neighborhood in the 75th percentile of the share of individuals with a college education and a neighborhood in the 25th percentile is only expected to be 3 percentage points, compared to the average effect of 13%. Similarly, the difference between neighborhoods in the 75% and 25% percentile in the median income, the share of blacks, the share of Asians and other races, and the share of Hispanics, is 5, 2, 1, and -5 percentage points, respectively.

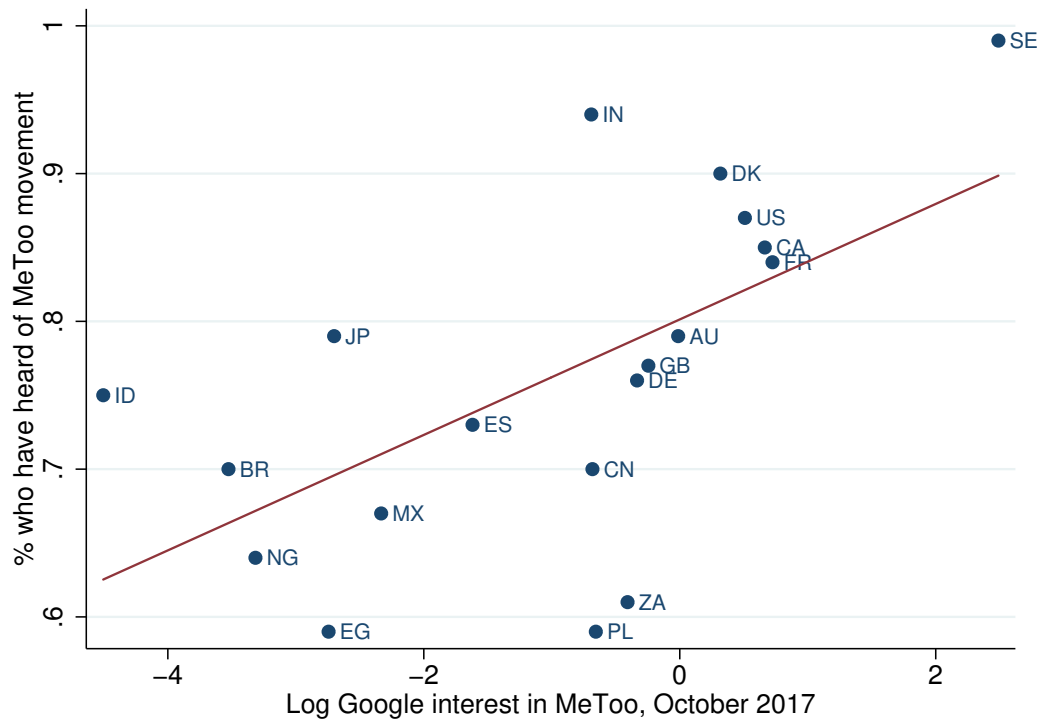
## **C Additional Figures and Tables**

Figure II.A.1: Newspaper Coverage



The first sub-figure shows the weekly average number of articles mentioning the term “metoo” in the newspapers USA Today, New York Post, Denver Post, and Chicago Sun-Times. The second sub-figure presents the weekly average number of articles mentioning the terms “sexual assault” or “sexual harassment” (articles mentioning both terms are counted twice). The vertical dashed line represents the start of the MeToo movement (October 2017). The newspapers were chosen based on circulation and data availability. The number of articles is determined using the website [newslibrary.com](http://newslibrary.com).

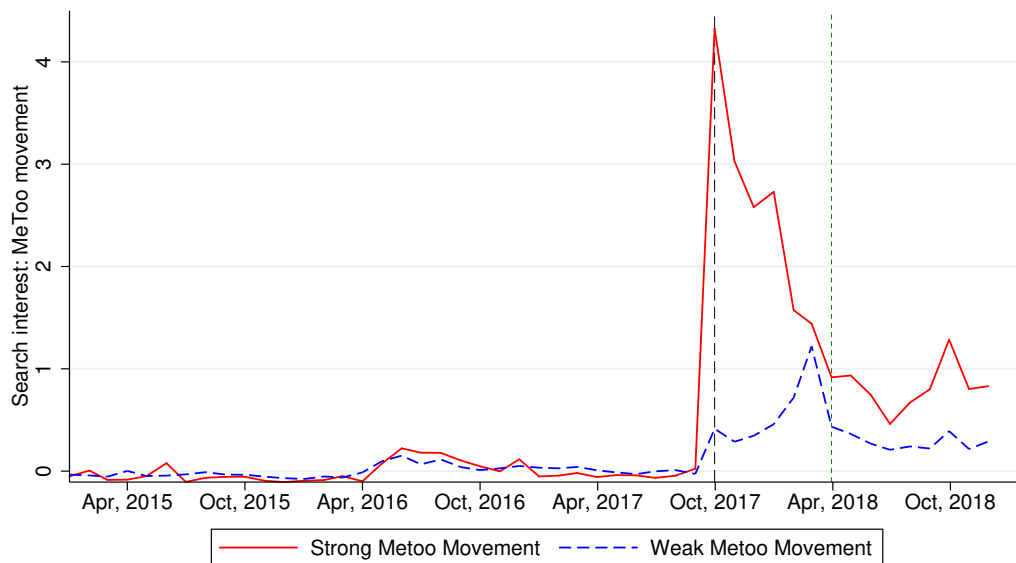
Figure II.A.2: Relationship Between Google Search Interest and Knowledge about the MeToo Movement



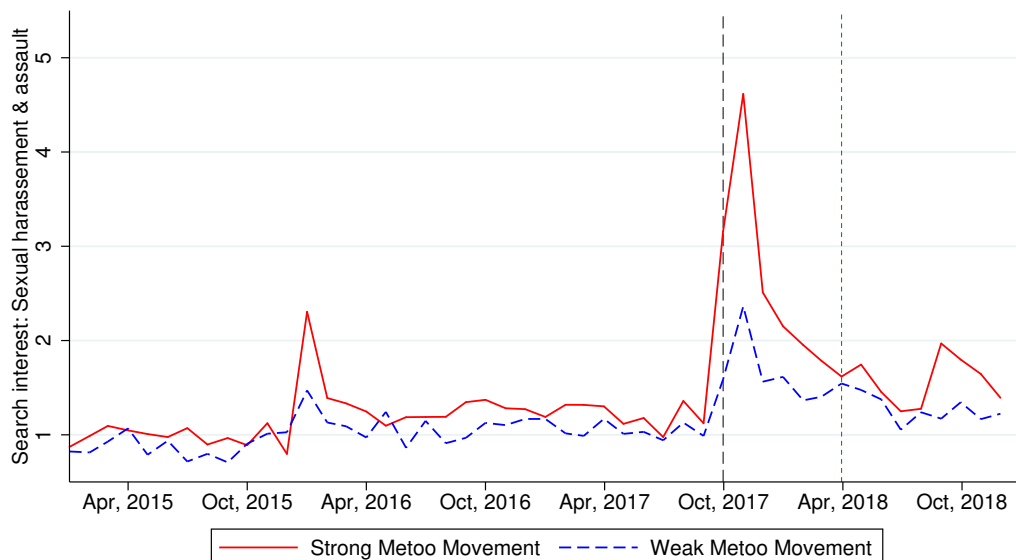
This figure shows the relationships between the log of Google search interest for terms related to the MeToo movement in October 2017 and the fraction of respondents who had heard about the MeToo movement in a YouGov survey conducted in February-March 2019 (YouGov, 2019).

Figure II.A.3: Search Interest by the Strength of the MeToo Movement

(a) Search Interest in the Topic of the MeToo Movement

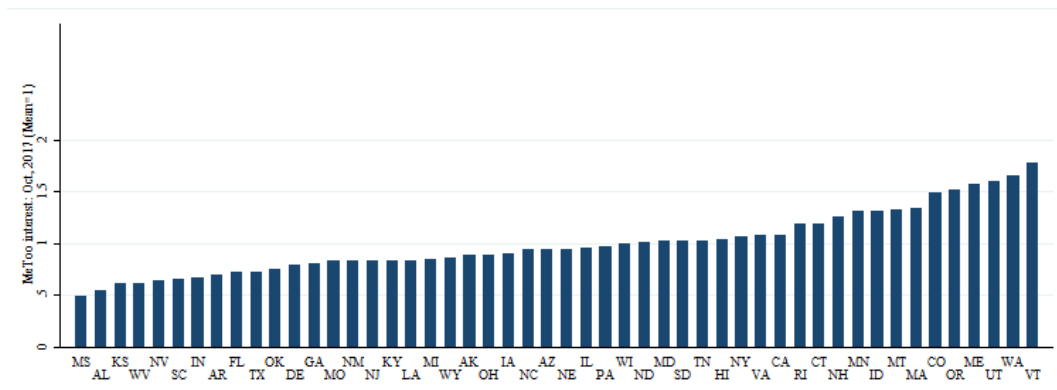


(b) Search interest in the Topics of Sexual Harassment and Sexual Assault



The figures show monthly search interest for OECD countries with strong and weak MeToo movements. Countries are classified as weak or strong by search interest in the MeToo topic in October 2017. Data is from Google Trends. The first vertical line represents the start of the MeToo movement, the second vertical line represents the end of the six month period we use to measure short-term effects. Sub-figure (a) shows search interest in the topic of the MeToo movement. Mean pre-MeToo interest is subtracted from the time series for each country separately so that the pre-MeToo period has a mean of zero, the data is then normalized so that the post-MeToo OECD mean equals 1. Sub-figure (b) shows search interest in the topics of sexual harassment and sexual assault. The data is normalized so that the pre-MeToo mean equals 1 for each country.

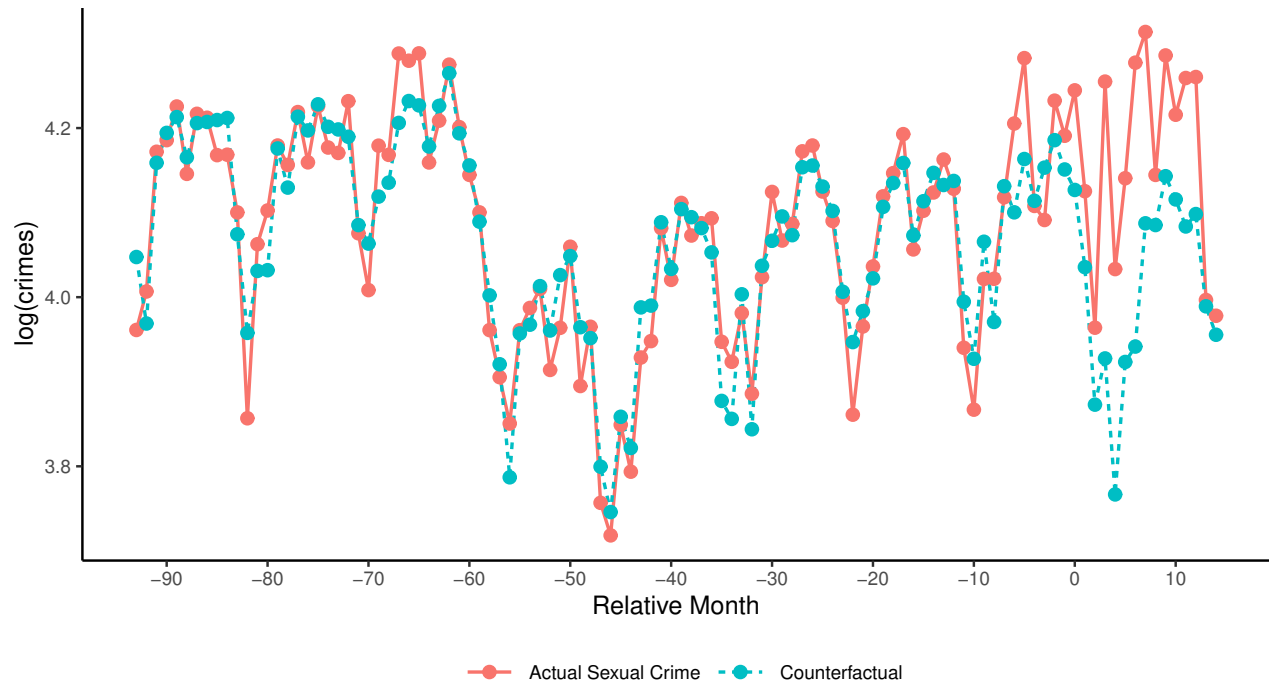
Figure II.A.4: Variation in MeToo Interest Across US States



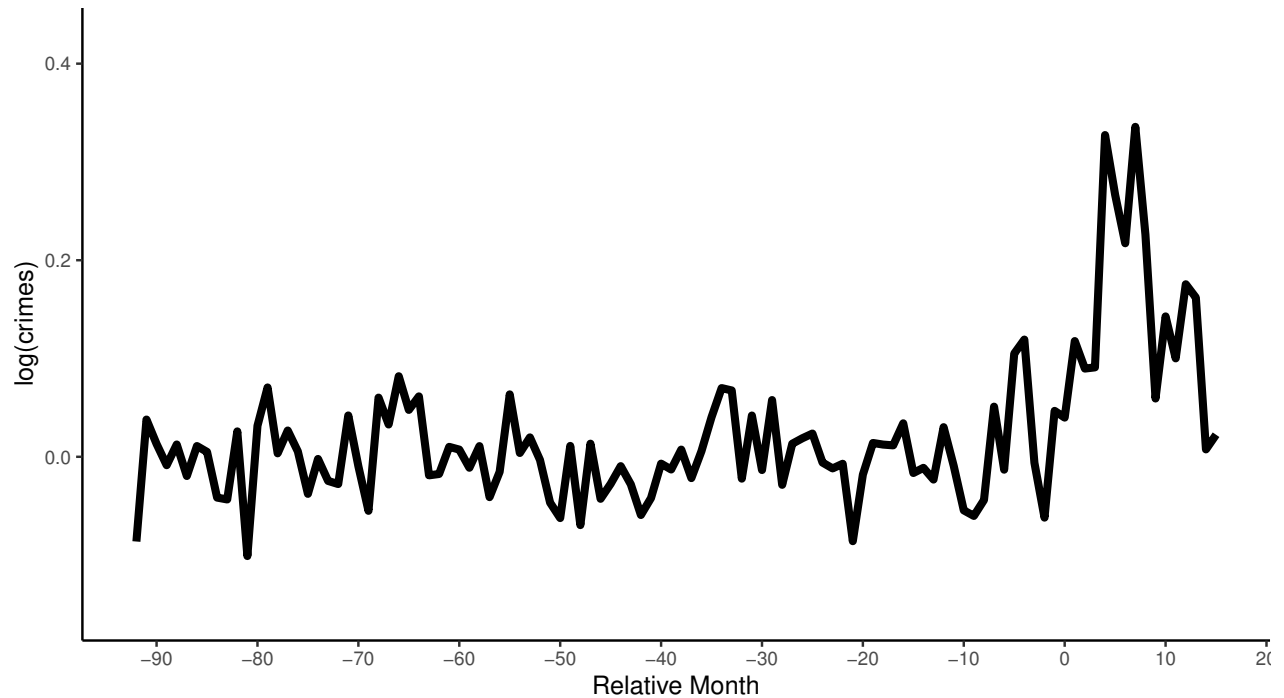
This figure shows the strength of the MeToo movement in US states, based on Google Search interest in the topic of the MeToo movement during October 2017.

Figure II.A.5: Matrix Completion Results

(a) Counterfactual Versus Actual Outcomes



(b) Average Treatment Effect



Sub-Figure (a) shows the actual and counterfactual reported sexual crimes (in logs) based on the matrix completion method for our sample of US cities. The method is described in Section 4.4. Sub-Figure (b) presents the average treatment effect - the difference between the actual crimes and the counterfactual. Standard errors are bootstrapped.

Table II.A.1: Effect of the MeToo Movement, Using Different MeToo Start Dates

	ln(crime)			
	(1)	(2)	(3)	(4)
Post MeToo start * Sexual Crime	0.092** (0.034)		0.083*** (0.030)	
Quarter of MeToo start * Sexual Crime		0.076* (0.043)		0.066** (0.028)
1Q after MeToo start * Sexual Crime		0.106** (0.049)		0.055 (0.045)
2Q after MeToo start * Sexual Crime		0.102** (0.041)		0.076** (0.033)
3Q after MeToo start * Sexual Crime		0.081** (0.036)		0.106*** (0.029)
4Q after MeToo start * Sexual Crime		0.090** (0.043)		0.103** (0.039)
Country * Crime type * Lin. trend	X	X	X	X
Country * Crime type * Quarter	X	X	X	X
Post MeToo start	X		X	
Quarters since MeToo start FE		X		X
Final quarter	Q4 2018	Q4 2018	Q4 2018	Q4 2018
Sample	MeToo only	MeToo only	MeToo only	MeToo only
Observations	1,276	1,276	1,324	1,324
Clusters	38	38	40	40
MeToo start indicator	MeToo search interest		SH/SA search interest	

This table shows the effect of the MeToo movement using different start dates of the MeToo movement in each country. In Columns (1) and (2) a start of the MeToo movement is the first month when searches for the MeToo movement topic was higher than the OECD median for October 2017. In Columns (3) and (4) a start of the MeToo movement is the first month, from October 2017 onward, when searches for the sexual harassment and sexual assault topics were the highest since 2010 in that country. Data from 30 OECD countries from 2010 to 2018. Standard errors clustered at the country by crime level in parenthesis. \*\*\*p<0.01, \*\*p<0.05, \*p<0.1.



Table II.A.2: MeToo Movement Start Date by Country

<b>Country</b>	<b>Start date using search interest in MeToo topic</b>	<b>Start date using search interest in sexual harassment and sexual assault topics</b>
Australia	October 2017	November 2017
Belgium	October 2017	No strong MeToo movement
Canada	October 2017	October, 2017
Chile	No strong MeToo movement	November 2017
Colombia	No strong MeToo movement	April 2018
Czech republic	November 2017	No strong MeToo movement
Denmark	October 2017	October 2017
Estonia	No strong MeToo movement	No strong MeToo movement
Finland	October 2017	October 2017
France	October 2017	October 2017
Greece	No strong MeToo movement	November 2017
Germany	October 2017	No strong MeToo movement
Iceland	October 2017	No strong MeToo movement
Ireland	October 2017	October 2017
Israel	No strong MeToo movement	November 2017
Japan	No strong MeToo movement	April 2018
Korea	February 2018	No strong MeToo movement
Lithuania	March 2018	November 2017
Mexico	No strong MeToo movement	November 2017
Netherlands	October 2017	No strong MeToo movement
New Zealand	October 2017	October 2017
Poland	No strong MeToo movement	No strong MeToo movement
Portugal	No strong MeToo movement	October 2017
Slovakia	No strong MeToo movement	No strong MeToo movement
Slovenia	No strong MeToo movement	December 2018
Switzerland	October 2017	October 2017
Spain	No strong MeToo movement	November 2017
Sweden	October 2017	October 2017
United Kingdom	October 2017	October 2017
United States	October 2017	October 2017

Table II.A.3: Definition of the Neighborhood Used by City

<b>City</b>	<b>Neighborhood Level</b>
Denver	Police District
Kansas City	Police Division
LA	Patrol Division
Louisville	Police Division
Nashville	MNPD Zone (Patrol Area)
New York City	Police Precinct
Seattle	Police Precinct

Table II.A.4: Effect of the MeToo Movement by Neighborhood

	ihts(crime)					
	(1)	(2)	(3)	(4)	(5)	(6)
Post * Sexual Crimes	0.128*** (0.020)	0.135*** (0.020)	0.128*** (0.020)	0.129*** (0.020)	0.129*** (0.020)	0.128*** (0.020)
Post * Sexual Crimes * Med. Income (std. dev.)		0.045** (0.020)				
Post * Sexual Crimes * % College			0.147 (0.096)			
Post * Sexual Crimes * % Blacks (Compared to Whites)				0.064 (0.093)		
Post * Sexual Crimes * % Other Race (Compared to Whites)					0.042 (0.132)	
Post * Sexual Crimes * % Hispanics						-0.148* (0.087)

Interquartile Range of Demographic		1.123	0.235	0.295	0.275	0.368
Diff. in Effect * 75th-25th Pct.		0.051	0.035	0.019	0.012	-0.055
Neighborhood * Crime Type * Lin. Trend	X	X	X	X	X	X
Neighborhood * Crime Type * Month	X	X	X	X	X	X
Post	X	X	X	X	X	X
Post * Demographic	X	X	X	X	X	X
Final Month	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18
Observations	25,056	25,056	25,056	25,056	25,056	25,056

This table shows the effect of the MeToo movement based on neighborhood-level data and tests for heterogeneous effects by neighborhood demographics. 2010-2018 city crime data. All regressions are weighted by the number of crimes that occurred in each neighborhood before the MeToo movement started. All demographic variables are first subtracted by their weighted mean. Robust standard errors in parenthesis. \*\*\*p<0.01, \*\*p<0.05; \*p<0.1

Table II.A.5: Data Sources for international\_data

Country	Data Providing Organization	Time period	Share of the population covered
Australia	New South Wales Bureau of Crime Statistics and Research, Queensland Police, Crime Statistics Agency of Victoria, and Western Australia Police	2010-2018	88%
Belgium	Federale politie	2010-2018	100%
Canada	Canadian Centre for Justice Statistics	2010-2018	100%
Chile	Policía de Investigaciones	2010-2018	100%
Colombia	Policía Nacional	2010-2018	100%
Czech republic	Policie České republiky	2010-2018	100%
Denmark	Danmarks Statistik	2010-2018	100%
Estonia	Politsei- ja Piirivalveamet	2010-2018	100%
Finland	Tilastokeskuksen	2010-2018	100%
France	Ministère de l'Intérieur	2010-2018	100%
Germany	Bundeskriminalamt	2012-2018	100%
Greece	Hellenic Statistical Authority (ELSTAT)	2010-2018	100%
Iceland	Ríkislögreglustjóri	2010-2018	100%
Ireland	Central Statistics Office	2010-2018	100%
Israel	Central Bureau of Statistics	2010-2018	100%
Japan	National Statistics Center	2015-2018	100%
Korea	Supreme prosecutors' office	2010-2018	100%
Lithuania	Informatikos ir Rysiu Departamentas	2012-2015 and 2017-2018	100%
Mexico	Instituto Nacional de Estadística y Geografía	2015-2018	100%
Netherlands	Korps Nationale Politie	2012-2018	100%
New Zealand	New Zealand Police	Q3 2014-2018	100%
Poland	Wydział ds. Parlamentarnych i Informacji Publicznej	2010-2018	100%
Portugal	Instituto Nacional de Estatística	2010-2018	100%
Slovakia	Statistický Úrad	2010-2018	100%
Slovenia	Statistični Urad	2010-2018	100%
Switzerland	Bundesamt für Statistik	2010-2018	100%
Spain	Ministerio del Interior	2010-2018	100%
Sweden	Brottsförebyggande rådet	2010-2018	100%
United Kingdom	Home Office: Crime and Policing Analysis Unit and Open Data Northern Ireland	2010-2018	92%
United States	Federal Bureau of Investigation	2010-2018	30%

Table II.A.6: Effect of the MeToo Movement in the US with Crime Aggregated by Offense Types

	lhs(crime)		
	(1)	(2)	(3)
Post * Sexual Assault	0.081*** (0.015)		
Post * Sexual Assault		0.096*** (0.017)	0.096*** (0.025)
State * Crime Type * Lin. Trend	X	X	X
State * Crime Type * Month	X	X	X
Post	X	X	X
Agg Crimes	Sexual/Other	NIBRS Categories	NIBRS Categories
S.E	Robust	Cluster by Crime Type	Cluster by Crime*State
Num of Clusters		21	735
Final Month	Mar 18	Mar 18	Mar 18
Observations	6,654	69,867	69,867

This table shows the effect of the MeToo movement using different crime aggregation and inference methods. Column (1) is our main estimate where crimes are categorized as either sexual crimes or non-sexual crimes, and robust standard errors are used. In columns (2)-(3), crimes are aggregated according to the NIBRS offense types. Incidents that include multiple offense types are excluded. Column (2) clusters standard errors by crime category and column (3) clusters by the interaction of state and crime category. All regressions are weighted by the number of crimes that occurred in each state before the MeToo movement started. 2010-2018 NIBRS data. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1.

Table II.A.7: Effect of the MeToo Movement by City

	(1)	(2)	(3)	ihs(crime)			(6)	(7)
Post * Sexual Crimes	0.144*** (0.041)	0.085*** (0.032)	0.189** (0.074)	0.083 (0.075)	0.401 (0.307)	-0.074 (0.082)	0.093 (0.065)	
Crime Type * Time	X	X	X	X	X	X	X	
Crime Type * Month	X	X	X	X	X	X	X	
Post	X	X	X	X	X	X	X	
Final Month	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18	
City	NYC	LA	Seattle	Denver	Nashville	Louisville	Kansas City	
Observations	198	198	198	126	126	198	198	

This table shows the effect of the MeToo movement on sexual crimes where the effect is calculated for each city separately. Robust standard error in parenthesis. \*\*\*p<0.01, \*\*p<0.05; \*p<0.1

Table II.A.8: Persistence of the Effect in the US

	ihs(crime)				
	(1)	(2)	(3)	(4)	(5)
Post * Sexual Crimes	0.100*** (0.011)		0.125*** (0.021)		
2017 Q4 * Sexual Crimes		0.070*** (0.017)		0.125*** (0.033)	0.113*** (0.039)
2018 Q1 * Sexual Crimes		0.093*** (0.020)		0.136** (0.065)	0.081 (0.067)
2018 Q2 * Sexual Crimes		0.101*** (0.018)		0.107*** (0.038)	0.090** (0.037)
2018 Q3 * Sexual Crimes		0.106*** (0.020)		0.138*** (0.035)	0.136*** (0.035)
2018 Q4 * Sexual Crimes		0.137*** (0.026)		0.115*** (0.038)	0.102** (0.041)
Location * Crime Type * Lin. Trend	X	X	X	X	X
Location * Crime Type * Month	X	X	X	X	X
Post	X	X	X	X	X
Data	NIBRS	NIBRS	Cities	Cities	Cities
Crimes	All	All	All	All	Reported Within 1 M
Observations	7,266	7,266	1,368	1,368	1,361

This table shows the effect of the MeToo movement on sexual crimes by quarter. Data is aggregated at the monthly state/city by crime category level. Columns (1) and (2) are based on 2010-2018 NIBRS data. Columns (3)-(5) are based on the sample of US cities. Columns (1) and (3) report the long-run effects until December 2018. Columns (2), (4), (5) report the effect by quarter. Column (5) includes only crimes that were reported within 30 days of their occurrence. Regressions are weighted by the number of crimes that occurred in each city before the MeToo movement started. Robust standard error in parenthesis. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

Table II.A.9: Effect of the MeToo Movement on Clearance

	ihb(crime)					
	(1)	(2)	(3)	(4)	(5)	(6)
Post * Sexual Assault, Not Cleared	0.106*** (0.016)			0.112*** (0.011)		
Post * Sexual Assault, Cleared	0.011 (0.024)			0.065*** (0.016)		
Post * Sexual Assault		0.025 (0.025)	0.103*** (0.017)		0.068*** (0.017)	0.115*** (0.011)
Difference	0.096***			0.047***		
State * Crime Type * Lin. Trend	X	X	X	X	X	X
State * Crime Type * Month	X	X	X	X	X	X
Post	X	X	X	X	X	X
Final Month	Mar 18	Mar 18	Mar 18	Dec 18	Dec 18	Dec 18
Crimes	All	Cleared	Not Cleared	All	Cleared	Not Cleared
Observations	9,981	6,654	6,654	10,899	7,266	7,266

This table shows the effect of the MeToo movement on sexual crimes by whether a case was cleared. A case is cleared if it has an arrest (a suspect is taken into custody based on a warrant or previously submitted report, arrested on view without a warrant or summoned to court), or if the police have sufficient probable cause to arrest a suspect but could not make an arrest for reasons outside their control including the victim refusing to cooperate, the death of the offender, the prosecutor declining prosecution for a reason other than lack of probable cause, the offender being in the custody of another jurisdiction, and the offender being a juvenile. In Column (1) and (4), the crimes are aggregated to three separate crime categories: Sexual crimes that were cleared, sexual crimes that were not cleared, and non-sexual crimes, which are the control group. In Columns (2) and (5), only crimes where the case was cleared are included and columns (3) and (6) include only crimes that were not cleared. Columns (1)-(3) focus on the short-run effect and columns (4)-(6) focus on the long-run effect. 2010-2018 NIBRS data. Regressions are weighted by the number of crimes that occurred in each state before the MeToo movement started. Robust standard errors in parenthesis. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1



Table II.A.10: Effect of Crime Covariates on Changes in the Sexual Assault Arrest Rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Post	-0.008*** (0.002)	-0.008*** (0.002)	-0.008*** (0.002)	-0.008*** (0.002)	-0.009*** (0.002)	-0.007*** (0.002)	-0.007*** (0.002)	-0.008*** (0.002)	-0.007*** (0.002)
Agency		X							X
Injury			X						X
Location				X					X
Relationship					X				X
Type						X			X
Weapon							X		X
Victim								X	X
Cal Month	X	X	X	X	X	X	X	X	X
Trend	X	X	X	X	X	X	X	X	X
Final Month	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18	Mar 18
Observations	625,172	625,172	625,172	625,172	625,172	625,172	625,172	625,172	625,172

This table shows the association between the post-period and arrests related to sexual assault when controlling for incident details. Each observation is a sexual assault crime reported between Jan 2010 and March 2018 and the outcome is whether the report resulted in an arrest. All columns control for a linear trend and calendar fixed effects. Column (1) shows that the arrest rate for sexual assault decreased in the post-period. Column (2)-(9) control for additional covariates. Column (2) control for agency fixed effects. Column (3) controls for whether the incidence results in an injury. Column (4) controls for the location type (residence, outside residence, unknown or multiple locations). Column (5) controls for the relationship between the victim and the offender (offender known to the victim, the offender is a stranger, unknown relationship or missing data). Column (7) controls for the type of sexual assault and whether the incident is associated with multiple offenses. Column (8) controls for the victim's race, sex, and age group. Column (9) controls for all the covariates. Robust standard errors in parenthesis. \*\*\*p<0.01; \*\*p<0.05; \*p<0.1

## References

- Acemoglu, D., T. A. Hassan, and A. Tahoun (2017). The power of the street: Evidence from Egypt's arab spring. *The Review of Financial Studies* 31(1), 1–42.
- Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics* 128(2), 469–530.
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics* 95(9-10), 1082–1095.
- Alsan, M. and M. Wanamaker (2017). Tuskegee and the health of black men. *The Quarterly Journal of Economics* 133(1), 407–455.
- Amenta, E., N. Caren, E. Chiarello, and Y. Su (2010). The political consequences of social movements. *Annual Review of Sociology* 36, 287–307.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association* 103(484), 1481–1495.
- Ashraf, A. (2019). Do performance ranks increase productivity? Evidence from a field experiment. Discussion Paper 196, Ludwig-Maximilians-Universität München und Humboldt-Universität zu Berlin.
- Ashraf, N., O. Bandiera, E. Davenport, and S. S. Lee (2020). Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *American Economic Review* 110(5), 1355–94.
- Ashraf, N., O. Bandiera, and B. K. Jack (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics* 120, 1–17.
- Athey, S. (2018). The impact of machine learning on economics. In *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.
- Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2017). Matrix Completion Methods for Causal Panel Data Models. Working paper.
- Bai, J., S. Jayachandran, E. J. Malesky, and B. A. Olken (2019). Firm growth and corruption: empirical evidence from vietnam. *The Economic Journal* 129(618), 651–677.
- Bandiera, O., I. Barankay, and I. Rasul (2010). Social incentives in the workplace. *The Review of Economic Studies* 77(2), 417–458.
- Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (2020). Improving police performance in Rajasthan, India: Experimental evidence on incentives, managerial autonomy and training. *American Economic Journal: Economic Policy*. Forthcoming.
- Banerjee, A., E. Duflo, C. Imbert, S. Mathew, and R. Pande (2020). E-governance, accountability, and leakage in public programs: Experimental evidence from a financial management reform in India. *American Economic Journal: Applied Economics* 12(4), 39–72.
- Banerjee, A., E. L. Ferrara, and V. Orozco (2019). Entertainment, education, and attitudes toward

- domestic violence. In *AEA Papers and Proceedings*, Volume 109, pp. 133–37.
- Banerjee, A. V. (1997). A theory of misgovernance. *The Quarterly Journal of Economics* 112(4), 1289–1332.
- Banerjee, A. V., E. Duflo, and R. Glennerster (2008). Putting a band-aid on a corpse: incentives for nurses in the Indian public health care system. *Journal of the European Economic Association* 6(2-3), 487–500.
- Bank, T. W. (2016). *World Development Report 2016: Digital Dividends*. World Bank Publications.
- Banuri, S. and P. Keefer (2013). Intrinsic motivation, effort and the call to public service. Policy Research Working Paper 6729, The World Bank.
- Barrera-Orsorio, F., K. Gonzalez, F. Lagos, and D. J. Deming (2020). Providing performance information in education: An experimental evaluation in colombia. *Journal of Public Economics* 186, 104185.
- Battaglini, M., R. B. Morton, and E. Patacchini (2020). Social groups and the effectiveness of protests. Working paper, National Bureau of Economic Research.
- Baumgartner, F. R. and C. Mahoney (2005). Social movements, the rise of new issues, and the public agenda. *Routing the opposition: Social movements, public policy, and democracy*, 65–86.
- Becker, G. S. and G. J. Stigler (1974). Law enforcement, malfeasance, and compensation of enforcers. *The Journal of Legal Studies* 3(1), 1–18.
- Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93(3), 491–507.
- Bertrand, M., R. Burgess, A. Chawla, and G. Xu (2020). The glittering prizes: Career incentives and bureaucrat performance. *The Review of Economic Studies* 87(2), 626–655.
- Bertrand, M., S. Djankov, R. Hanna, and S. Mullainathan (2007). Obtaining a driver’s license in india: an experimental approach to studying corruption. *The Quarterly Journal of Economics* 122(4), 1639–1676.
- Bertrand, M., E. Kamenica, and J. Pan (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics* 130(2), 571–614.
- Bhatnagar, A., A. Mathur, A. Munasib, and D. Roy (2019). Sparking the #MeToo revolution in India: The ‘Nirbhaya’ case in Delhi. Working paper, American Enterprise Institute.
- Blader, S., C. Gartenberg, and A. Prat (2020). The contingent effect of management practices. *The Review of Economic Studies* 87(2), 721–749.
- Bø, E. E., J. Slemrod, and T. O. Thoresen (2015). Taxes on the internet: Deterrence effects of public disclosure. *American Economic Journal: Economic Policy* 7(1), 36–62.
- Bold, T., M. Kimenyi, G. Mwabu, A. Ng’ang’a, and J. Sandefur (2018). Experimental evidence on scaling up education reforms in kenya. *Journal of Public Economics* 168, 1–20.
- Bottan, N. L. and R. Perez-Truglia (2015). Losing my religion: The effects of religious scandals on religious participation and charitable giving. *Journal of Public Economics* 129, 106–119.

- Burstein, P. and S. Sausner (2005). The incidence and impact of policy-oriented collective action: competing views. *Sociological Forum* 20(3), 403–419.
- Bursztyn, L., G. Egorov, and S. Fiorin (2017). From extreme to mainstream: How social norms unravel. Working paper, National Bureau of Economic Research.
- Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2018). Misperceived social norms: Female labor force participation in Saudi Arabia. Working paper, National Bureau of Economic Research.
- Callen, M., S. Gulzar, A. Hasanain, M. Y. Khan, and A. Rezaee (2020). Data and policy decisions: Experimental evidence from pakistan. *Journal of Development Economics* 146, 102523.
- Caputi, T. L., A. L. Nobles, and J. W. Ayers (2019). Internet searches for sexual harassment and assault, reporting, and training since the #MeToo movement. *JAMA internal medicine* 179(2), 258–259.
- Castle, J. J., S. Jenkins, C. D. Ortals, L. Poloni-Staudinger, and J. C. Strachan (2020). The effect of the #metoo movement on political engagement and ambition in 2018. *Political Research Quarterly* 73(4), 926–941.
- Charles, K. K., J. Guryan, and J. Pan (2018). The effects of sexism on american women: The role of norms vs. discrimination. Working paper, National Bureau of Economic Research.
- Cheng, I.-H. and A. Hsiaw (2019). Reporting sexual misconduct in the #MeToo era. Working paper.
- Chong, A. and E. L. Ferrara (2009). Television and divorce: Evidence from Brazilian novelas. *Journal of the European Economic Association* 7(2-3), 458–468.
- Christensen, D. and F. Garfias (2018). Can you hear me now? How communication technology affects protest and repression. *Quarterly journal of political science* 13(1), 89.
- Clot, S., G. Grolleau, and L. Ibanez (2018). Moral self-licencing and social dilemmas: an experimental analysis from a taking game in madagascar. *Applied Economics* 50(27), 2980–2991.
- Cornelissen, T., C. Dustmann, and U. Schönberg (2017). Peer effects in the workplace. *American Economic Review* 107(2), 425–56.
- Cowley, E. and S. Smith (2014). Motivation and mission in the public sector: Evidence from the world values survey. *Theory and Decision* 76(2), 241–263.
- Dal Bó, E., F. Finan, N. Y. Li, and L. Schechter (2019). Government decentralization under changing state capacity: Experimental evidence from paraguay. Working Paper 24879, National Bureau of Economic Research.
- Dhar, D., T. Jain, and S. Jayachandran (2018). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india.
- Djankov, S., C. Freund, and C. S. Pham (2010). Trading on time. *The Review of Economics and Statistics* 92(1), 166–173.
- Djankov, S., D. Georgieva, and R. Ramalho (2018). Business regulations and poverty. *Economics Letters* 165, 82–87.
- Djankov, S., C. McLiesh, and R. M. Ramalho (2006). Regulation and growth. *Economics Letters* 92(3), 395–401.

- Dodge, E., Y. Neggers, R. Pande, and C. T. Moore (2018). Having it at hand: How small search frictions impact bureaucratic efficiency. Working paper.
- Dustan, A., S. Maldonado, and J. M. Hernandez-Agramonte (2018). Motivating bureaucrats with non-monetary incentives when state capacity is weak: Evidence from large-scale field experiments in peru. Working Paper 136, Peruvian Economic Association.
- Enikolopov, R., A. Makarin, and M. Petrova (2019). Social media and protest participation: Evidence from Russia. Working paper, SSRN.
- Finan, F., B. A. Olken, and R. Pande (2017). The personnel economics of the developing state. In *Handbook of Economic Field Experiments*, Volume 2, pp. 467–514. Elsevier.
- Folke, O., J. Rickne, S. Tanaka, and Y. Tateishi (2020). Sexual harassment of women leaders. *Daedalus* 149(1), 180–197.
- Freund, C., M. Hallward-Driemeier, and B. Rijkers (2016). Deals and delays: Firm-level evidence on corruption and policy implementation times. *World Bank Economic Review* 30(2), 354–382.
- García-Jimeno, C., A. Iglesias, and P. Yildirim (2018). Women, rails and telegraphs: An empirical study of information diffusion and collective action. Working paper, National Bureau of Economic Research.
- Gerber, A. S., D. P. Green, and C. W. Larimer (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, 33–48.
- Green, D. P., A. M. Wilke, and J. Cooper (2020). Countering violence against women by encouraging disclosure: A mass media experiment in rural uganda. *Comparative Political Studies* 53(14), 2283–2320.
- Guriev, S. (2004). Red tape and corruption. *Journal of development economics* 73(2), 489–504.
- Hague Institute for Innovation of Law (2018). Justice needs and satisfaction in Bangladesh. Research report, Hague Institute for Innovation of Law.
- Hersch, J. (2011). Compensating differentials for sexual harassment. *American Economic Review* 101(3), 630–34.
- Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, 74–91.
- Huntington, S. P. (1968). *Political Order in Changing Societies*. New Haven: Yale University Press.
- International Lawyers Network (2019). Sexual harassment in the workplace: What employers need to know. Report.
- Ipsos (2017a). Ipsos/NPR Examine How Views on Sexual Harassment Have Changed in the Past Year.
- Ipsos (2017b). The #MeToo Movement: One Year Later.
- Iyer, L., A. Mani, P. Mishra, and P. Topalova (2012). The power of political voice: women’s political representation and crime in India. *American Economic Journal: Applied Economics* 4(4), 165–193.
- Jayaraman, R., D. Ray, and F. De Véricourt (2016). Anatomy of a contract change. *American Economic Review* 106(2), 316–58.

- Jensen, R. and E. Oster (2009). The power of TV: Cable television and women's status in India. *The Quarterly Journal of Economics* 124(3), 1057–1094.
- Kaufmann, D. and S.-J. Wei (1999). Does "grease money" speed up the wheels of commerce? Working Paper 7093, National Bureau of Economic Research.
- Khan, A. Q., A. I. Khwaja, and B. A. Olken (2016). Tax farming redux: Experimental evidence on performance pay for tax collectors. *The Quarterly Journal of Economics* 131(1), 219–271.
- Khan, A. Q., A. I. Khwaja, and B. A. Olken (2019). Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings. *American Economic Review* 109(1), 237–70.
- Klapper, L., L. Laeven, and R. Rajan (2006). Entry regulation as a barrier to entrepreneurship. *Journal of Financial Economics* 82(3), 591–629.
- La Ferrara, E., A. Chong, and S. Duryea (2012). Soap operas and fertility: Evidence from Brazil. *American Economic Journal: Applied Economics* 4(4), 1–31.
- Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies* 76(3), 1071–1102.
- Leff, N. H. (1964). Economic development through bureaucratic corruption. *American Behavioral Scientist* 8(3), 8–14.
- Lewis-Faupel, S., Y. Neggers, B. A. Olken, and R. Pande (2016). Can electronic procurement improve infrastructure provision? Evidence from public works in India and Indonesia. *American Economic Journal: Economic Policy* 8(3), 258–83.
- Lins, K. V., L. Roth, H. Servaes, and A. Tamayo (2020). Gender, Culture, and Firm Value: Evidence from the Harvey Weinstein Scandal and the# MeToo Movement.
- Lonsway, K. A. and J. Archambault (2012). The "justice gap" for sexual assault cases: Future directions for research and reform. *Violence against women* 18(2), 145–168.
- Luo, H. and L. Zhang (2020). Scandal, social movement, and change: Evidence from# metoo in hollywood.
- Madestam, A., D. Shoag, S. Veuger, and D. Yanagizawa-Drott (2013). Do political protests matter? Evidence from the tea party movement. *The Quarterly Journal of Economics* 128(4), 1633–1685.
- Martinez-Bravo, M. and A. Stegmann (2018). In vaccines we trust? The effects of the CIA's vaccine ruse on immunization in Pakistan. Working paper.
- Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review* 99(1), 112–45.
- Maskin, E. and J. Riley (1984). Monopoly with incomplete information. *The RAND Journal of Economics* 15(2), 171–196.
- McDougal, L., S. Krumholz, N. Bhan, P. Bharadwaj, and A. Raj (2018). Releasing the tide: How has a shock to the acceptability of gender-based sexual violence affected rape reporting to police in India? *Journal of interpersonal violence*, 1–23.
- Miller, A. R. and C. Segal (2019). Do female officers improve law enforcement quality? effects on crime reporting and domestic violence. *The Review of Economic Studies* 86(5), 2220–2247.

- MIT Election Data and Science Lab (2018). County Presidential Election Returns 2000-2016.
- Morabito, M. S., L. M. Williams, and A. Pattavina (2019). Decision making in sexual assault cases: Replication research on sexual violence case attrition in the US. Report, US Department of Justice, National Institute of Justice NCJ.
- Muralidharan, K. and P. Niehaus (2017). Experimentation at scale. *Journal of Economic Perspectives* 31(4), 103–24.
- Muralidharan, K., P. Niehaus, S. Sukhtankar, and J. Weaver (2020). Improving last-mile service delivery using phone-based monitoring. *American Economic Journal: Applied Economics*. Forthcoming.
- Mussa, M. and S. Rosen (1978). Monopoly and product quality. *Journal of Economic Theory* 18(2), 301–317.
- Myrdal, G. (1968). *Asian drama, an inquiry into the poverty of nations*. London: The Penguin Press.
- Nicoletti, G. and S. Scarpetta (2003). Regulation, productivity and growth: OECD evidence. *Economic Policy* 18(36), 9–72.
- Niehaus, P. and S. Sukhtankar (2013a). Corruption dynamics: The golden goose effect. *American Economic Journal: Economic Policy* 5(4), 230–69.
- Niehaus, P. and S. Sukhtankar (2013b). The marginal rate of corruption in public programs: Evidence from India. *Journal of Public Economics* 104, 52–64.
- Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in indonesia. *Journal of Political Economy* 115(2), 200–249.
- Onwuachi-Willig, A. (2018). What about #UsToo: The invisibility of race in the #MeToo movement. *Yale LJF* 128, 105.
- Perez-Truglia, R. and U. Troiano (2018). Shaming tax delinquents. *Journal of Public Economics* 167, 120–137.
- Pew Research Center (2018). Activism in the social media age. Research report.
- Rasul, I. and D. Rogger (2018). Management of bureaucrats and public service delivery: Evidence from the nigerian civil service. *The Economic Journal* 128(608), 413–446.
- Reinikka, R. and J. Svensson (2005). Fighting corruption to improve schooling: Evidence from a newspaper campaign in uganda. *Journal of the European Economic Association* 3(2-3), 259–267.
- Rose-Ackerman, S. (1978). *Corruption: A study in political economy*. New York: Academic Press.
- Rosenzweig, M. R. and C. Udry (2020). External validity in a stochastic world: Evidence from low-income countries. *The Review of Economic Studies* 87(1), 343–381.
- Rotenberg, C. and A. Cotter (2018). Police-reported sexual assaults in Canada before and after #MeToo, 2016 and 2017. *Juristat: Canadian Centre for Justice Statistics*, 1–27.
- Sachdeva, S., R. Iliev, and D. L. Medin (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological science* 20(4), 523–528.
- Singh, A. (2020). Myths of official measurement: Auditing and improving administrative data in

- developing countries. Working Paper 20/042, RISE.
- Spohn, C. and K. Tellis (2012). Policing and prosecuting sexual assault in Los Angeles city and county: A collaborative study in partnership with the Los Angeles Police Department, the Los Angeles County Sheriff's Department, and the Los Angeles County District Attorney's Office. Report, US Department of Justice, National Institute of Justice NCJ.
- Sunstein, C. R. (2019). *How Change Happens*. MIT Press.
- Svensson, J. (2003). Who must pay bribes and how much? Evidence from a cross section of firms. *The Quarterly Journal of Economics* 118(1), 207–230.
- Transparency International Bangladesh (2016, 06). Corruption in service sectors, national household survey 2015. Research report, Transparency International Bangladesh.
- Transparency International Bangladesh (2018). Corruption in service sectors: National household survey 2017. Research report, Transparency International Bangladesh.
- Wasow, O. (2020). Agenda seeding: How 1960s black protests moved elites, public opinion and voting. *American Political Science Review*, 1–22.
- Weaver, J. (2020). Jobs for sale: Corruption and misallocation in hiring. Working paper.
- YouGov (2019). Metoo favourability. Survey results, YouGov-Cambridge Globalism Project.



ProQuest Number: 28317783

INFORMATION TO ALL USERS

The quality and completeness of this reproduction is dependent on the quality and completeness of the copy made available to ProQuest.



Distributed by ProQuest LLC (2021).

Copyright of the Dissertation is held by the Author unless otherwise noted.

This work may be used in accordance with the terms of the Creative Commons license or other rights statement, as indicated in the copyright statement or in the metadata associated with this work. Unless otherwise specified in the copyright statement or the metadata, all rights are reserved by the copyright holder.

This work is protected against unauthorized copying under Title 17,  
United States Code and other applicable copyright laws.

Microform Edition where available © ProQuest LLC. No reproduction or digitization of the Microform Edition is authorized without permission of ProQuest LLC.

ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 - 1346 USA